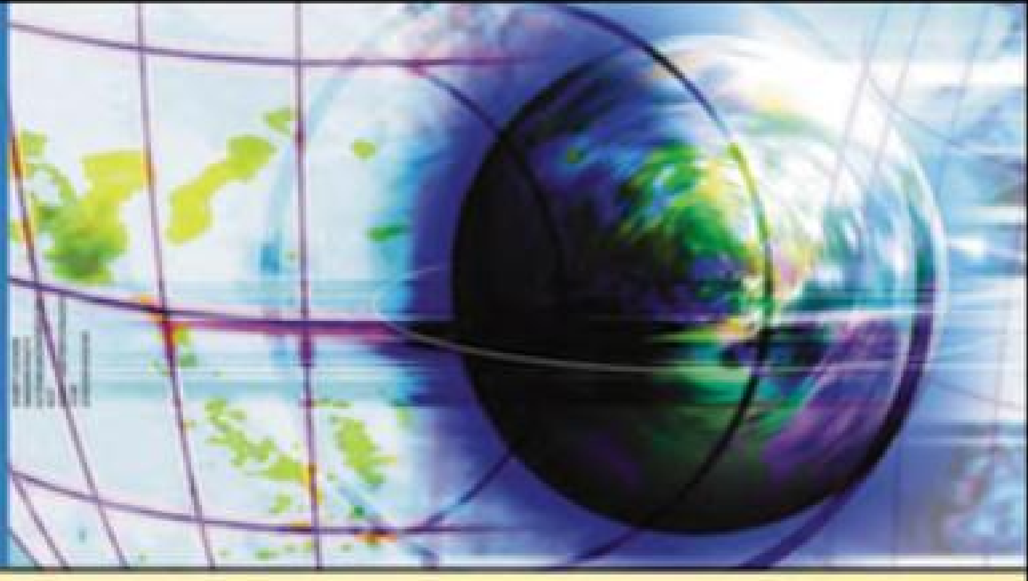
WILEY HANDBOOKS IN OPERATIONS RESEARCH AND MANAGEMENT SCIENCE



HANDBOOK OF INTEGRATED RISK MANAGEMENT IN GLOBAL SUPPLY CHAINS

EDITED BY

Panos Kouvelis
Lingxiu Dong
Onur Boyabatlı
Rong Li

WILEY

THE HANDBOOK OF

Integrated Risk Management in Global Supply Chains

Wiley Handbooks in

OPERATIONS RESEARCH AND MANAGEMENT SCIENCE

The *Wiley Handbooks in Operations Research and Management Science* is a series of books that present both established techniques and cutting-edge developments in the field of operations research and management science. The goal of each handbook is to supply a practical, one-stop reference that treats the theory, formulae, and applications that, together, make up the cornerstones of a particular topic in the field. A self-contained presentation allows each volume to serve as a quick reference on ideas and methods for practitioners, while providing an accessible introduction to key concepts for students. The result is a high-quality, comprehensive collection that serves as a mainstay for novices and professionals alike.

Kouvelis, Dong, Boyalatı, and Li · *Handbook of Integrated Risk Management in Global Supply Chains*

Forthcoming *Wiley Handbooks in Operations Research and Management Science*

Sokolowski and Banks · *Handbook of Real-World Applications of Modeling and Simulation*

The Handbook of

Integrated Risk Management in Global Supply Chains

Edited by

Panos Kouvelis
Lingxiu Dong
Onur Boyabatli
Rong Li

WILEY

A JOHN WILEY & SONS, INC., PUBLICATION

Published by John Wiley & Sons, Inc., Hoboken, New Jersey
Published simultaneously in Canada

Library of Congress Cataloging-in-Publication Data

Handbook of integrated risk management in global supply chains / Panos Kouvelis. . . [et al.].
p. cm. – (Wiley handbooks in operations research and management science)
Includes bibliographical references and index.
ISBN 978-0-470-53512-7
1. Business logistics. 2. Delivery of goods–Management. 3. Risk management. 4. Globalization. I. Kouvelis, Panos.
HD38.5 .H354 658.5–dc22 2011015212

Printed in the United States of America

oBook ISBN: 978-1-118-11580-0
ePDF ISBN: 978-1-118-11577-0
ePub ISBN: 978-1-118-11579-4
eMobi ISBN: 978-1-118-11578-7

10 9 8 7 6 5 4 3 2 1

Contents

PART TWO
Integrated Risk Management: Operations and Finance Interface

Foreword

Globalization and the unbundling of value chains have been the major factors in the growth of global supply chains in the past two decades. These reflect the two fundamental factors driving economic growth identified over two centuries ago by Adam Smith in *The Wealth of Nations*: namely specialization (to achieve economies of scale) and trade (to link the most cost-effective sources of product design and manufacturing to end markets). Coupled with trade liberalization and the benefits of Internet-based IT, unbundling and globalization of trade have driven a veritable explosion of economic growth in the 1990s to the present. This has included the developments in market-based and financial institutions of the European Union and the increasing salience of the BRIC countries (Brazil, Russia, India, and China). In particular, China and India began their ascent to global leadership in low-cost manufacturing and services, including information-based technology support. All of this has been reflected in the huge increases in outsourcing and offshoring evident in the past decade for low-cost sources of goods and services and the unbundling of global supply chains.

While business and entrepreneurship, and rational trade policy, are the natural vehicles for realizing the power of globalization, operations and global fulfillment architectures have become the primary "glue" for integrating multitiered, global networks. The "new operations" that have emerged in the past decade reflects a strategic view of the supply chain and greater emphasis on information and financial flows across the network. Starting with Michael Porter's work on the value chain in the 1980s, and motivated by the huge success of the supply-chain rationalizations of the 1990s, many of the most successful and innovative companies have come to formulate their strategies and business models in operational terms. These have come in the guise of innovative approaches to supply chain design based on operational flexibility and the network reconfiguration and sourcing strategies. In this evolution, companies have moved from a narrow focus on cost and leanness to an appreciation of the customer's willingness to pay for reliability and tailored logistics solutions and to a closer scrutiny of the total financial costs and risks of supply relationships. Against this background, this text is a timely contribution to the critical and developing theme of integrated risk management for supply strategy. With the noted increasing organizational and geographic complexity of supply operations, business leaders face a complex fabric of risks from extreme weather events to major accidents to financial crises, in addition to the normal business risks of coordinating supply and demand through effective supply operations. Understanding and mitigating these risks by all supply chain participants is now recognized as an essential accompaniment to profitability and long-term value creation.

The editors have brought together here the front line of the management science research community dealing with these issues. They have confronted them with a double challenge: first, to summarize the frontiers of research on integrated risk management for supply chain design and operations, and second, to draw the implications of this research for practice. The result is a splendid synthesis and contribution to our knowledge of how global supply chains are evolving and the fundamental role of risk management in assuring their robustness and resilience. In the process, the papers here also indicate the importance of integrating operations and finance in assuring profitability in the networked environment, which is now the essential frame for companies and economies across the globe. It is a distinct pleasure to see this set of essays appear, mapping both our current state of knowledge and the challenges ahead for business and research.

PAUL R. KLEINDORFER

Anheuser-Busch Professor of Management Science, Emeritus
The Wharton School of the University of Pennsylvania, and
Paul Dubrule Professor of Sustainable Development, INSEAD

Preface

Risks have always been a part of business reality. However, the extraordinary growth of global trade in the past several decades have taken the multitude and magnitude of risks to a new level of significance.The trend of globalization has put virtually every company in the massive, complex global trade network. On the one hand, the interconnection introduces to supply chains tremendous opportunity of cost reduction, access to labor/talent pools, capitals, and markets. On the other hand, the increase in supply chain scope presents new challenges that go beyond the typical supply chain concern of demand and supply uncertainties. Companies that were traditionally local or national oriented are now affected by the world commodity price shocks, currency fluctuations, and even by trade policy and law changes in other countries. Conversely, glitches at a supplier in one country may cause significant business disruptions downstream in another country. Added to the complexity in supply chain paradigm and convoluted dependence among supply chain members is the increasing frequency of natural disasters (hurricane, earthquake, volcano eruption, etc.) and political instability in parts of the world, which consequently superimpose another layer of change to companies' risk profiles. Local catastrophic events more often than not have impacts on businesses around the world, and collectively they are no longer rare events that companies can afford to merely react to on an ad hoc basis.

Companies have come under growing pressure to develop risk management schemes that identify the risks they are facing, measure the likelihood of occurrence and the scope and magnitude of the impact, and mitigate the detrimental impact. The urgency of seeking comprehensive risk management tools has motivated academicians to expand the supply chain management research in the direction of incorporating various aforementioned risks into the scope of study. Recognizing the multifunctional nature of the challenges faced by business, this growing area of research, appropriately titled as "integrated risk management," fully embraces a multidisciplinary approach that exploits recent development in finance, decision theory, operations research, and supply chain management to manage the complex, highly interacting, and diverse global supply chain risks.

The main objectives of this handbook are

1. To provide a collection of original ideas on the integrated risk management in operations and global supply chain to help academicians and practitioners develop a multifunctional perspective on these issues.

2. To offer managerial insights and outline challenges in identifying, measuring, and controlling risks in practice.
3. To raise important questions that remain unanswered, so as to set the agenda for future research that fully reflects the interdisciplinary nature of the risk management, measurement, and control challenges in operations and global supply chains.

We have organized 20 chapters of the handbook in five Parts. Part I, Foundation and Overview, provides general frameworks in supply chain risk management and overviews of some theoretical methodology. Parts II, III, IV, are devoted to the three themes of research on risk mitigation strategies, respectively, Integrated Risk Management, Supply Chain Finance, and Operational Risk Management Strategies. Lastly, Part V, Industrial Applications, showcases supply chain risk management in practice.

Foundation and Overview

Part I consists of five chapters. In Chapter 1, Integrated Risk Management: A Conceptual Framework with Research Overview and Applications in Practice, Kouvelis, Dong, Boyabatlı, and Li present an action-based supply chain risk management framework with insightful discussions on categorizations of risks and on actions should be taken in the planning and execution stages of risk management. Mapping the growing research field of integrated risk management within the context of the proposed conceptual framework, they identify a number of interesting directions for future research.

In Chapter 2, Risk Management and Operational Hedging: An Overview, Jan Van Mieghem provides an introduction and overview of risk management and broad operational hedging techniques, offers a view of treating risk management as an integral part of operations strategy, and describes in detail a four-step risk management process including risk identification, risk assessment, tactical risk decisions, and operational hedging implementation. The author also briefly discusses the financial hedging of operational risks.

The sense of urgency that the need for disruption risk management tools is universal across industries is clearly communicated in Chapter 3, The Effect of Supply Chain Disruptions on Corporate Performance, by Kevin Hendricks and Vinod Singhal. They investigate the impact of supply chain disruptions on the corporate performance. The evidence indicates that firms continue to operate at a lower performance level for at least two years after experiencing disruptions. Given the significant economic losses, firms cannot afford such disruptions even if they occur infrequently. A key managerial insight is that overemphasis on efficiency and removing slack from the system can make supply chains vulnerable, unreliable, and nonresponsive: While efficient and lean supply chains are desirable objectives, they should not come at the expense of reliability and responsiveness.

Different operational strategies have different strengths and limitations in managing supply chain disruption risks. In Chapter 4, Operational Strategies for Managing Supply Chain Disruption Risk, Brian Tomlin and Yimin Wang provide a roadmap to supply chain managers in designing and implementing their disruption mitigation strategy, including stockpiling inventory, supply diversification, backup supply, demand management, and supply chain strengthening. The key managerial insight is that a one-size-fits-all approach of employing the same strategy for all product lines may not be appropriate if different products exhibit different supply chain and market characteristics. The disruption management strategy should be tailored to the needs of each product.

When firms do not have sufficient data to characterize the probability distribution of risks they are facing, they are exposed to decision making under ambiguity. Using traditional performance measures such as expected value or mean-variance utility becomes problematic, and there is a need for defining appropriate performance measures under ambiguity. Karthik Natarajan, Melvyn Sim, Chung-Piaw Teo, in Chapter 5, Beyond Risk: Ambiguity in Supply Chains, address this issue. They review the notion of ambiguity (origination, evidence, and models) from different academic fields and link it to supply chain management practice. They discuss one of the most popular approaches to account for aversion to ambiguity, the maximin expected utility (MEU) theory, and demonstrate the implications of this theory in making operational decisions in a supply chain framework.

Integrated Risk Management

Part II, Integrated Risk Management, consists of four chapters, all exploring the joint use of operational and financial hedging of commodity price uncertainties.

In Chapter 6, Managing Storable Commodity Risks: Role of Inventories and Financial Hedges, Panos Kouvelis, Rong Li, and Qing Ding consider a risk-averse buyer who dynamically maximizes the mean-variance utility of the cash flows in a multiperiod setting. The commodity buyer procures from a long-term supply contract and from the commodity market, and hedges the price and demand uncertainties dynamically, using all financial derivatives written on the commodity price, such as futures contracts, call and put options, and so on. They derive an optimal time-consistent integrated policy of inventory and financial hedging. They find, contrary to results in the existing literature, that myopic hedging is not optimal; financial hedging may lead to inventory reduction in multiperiod problems. Finally, insights are offered on the impact of the physical inventory and financial hedging on the profitability, variance control, and service level to the customer.

In Chapter 7, Integrated Production and Risk Hedging with Financial Instruments, Çağri Haksöz and Sridhar Seshadri provide a review of the existing literature on integrated operational hedging and financial hedging decisions for a risk-averse firm that produces a commodity and maximizes the expected utility (general utility or mean-variance objective), in single-period and multiperiod settings. They investigate the effectiveness of the optimal use of a fixed-price long-term contract and spot market trading for a risk neutral commodity producer in a continuous-time model.

Contingency planning is a powerful risk-mitigating tool and is widely used in industry. Bardia Kamrad, in Chapter 8, Capacity Expansion as A Contingent Claim: Flexibility and Real Options in Operations, provides a review of the real options framework for evaluating contingency claims. In this methodological chapter, he discusses the no-arbitrage financial pricing model for contingent claims (e.g., options) and introduces how to apply this model to value real options in the context of a capacity investment problem. Advantages and disadvantages, rising from the difference between financial options and real options, are also discussed.

In Chapter 9, Financial Valuation of Supply Chain Contracts, Mustafa Pınar, AlperŞen, and A. Gökay Erön apply the no-arbitrage financial pricing model for contingent claims to value multiperiod supply contracts. They consider a setting where the demand is perfectly positively correlated with a risky security, and both the financial and the demand markets evolve as discrete scenario trees. The buyer is committed at the beginning of the planning horizon to purchase a fixed quantity of inventory each period and a multiperiod option contract, which allows the buyer to determine how much to buy (or how many options to exercise) each period after observing real demand. An experimental study is presented to illustrate the sensitivity analysis.

Supply Chain Finance

Part III, Supply Chain Finance, consists of four chapters. The first two chapters study financing alternatives available to budget-constrained supply chain members in a linear supply chain of a newsvendor retailer and a supplier, as well as the presence of a financial institution (e.g., a bank). In Chapter 10, Supply Chain Finance, Panos Kouvelis and Wenhui Zhao consider supply contracts that could effectively serve as trade credit contracts that the supplier offers to the retailer. A key managerial insight is that when both the supplier and the retailer have access to competitively priced short-term bank loans, it is in the best interest of the risk neutral supplier to offer trade credit to the retailer with an interest rate lower than or equal to the risk free rate; and if optimally parameterized, the retailer always prefer such trade credit. Contrasting to the known result that buyback contracts, revenue sharing contracts, and all-unit discount contracts all coordinate a supply chain without budget constraints, the authors demonstrate that none of these contracts coordinate a budget-constrained supply chain though yield different levels of supply chain efficiency.

In Chapter 11, The Role of Financial Services in Procurement Contracts, René Caldentey and Xiangfeng Chen consider a model that is similar to the previous chapter except a difference in the assumption of when the retailer's budget constraint limits his ordering quantity. The authors reach similar conclusions regarding the benefit of trade credit over bank loans. A key managerial insight is that the value of trade-credit for the retailer is non-monotonic in his internal budget, and there is an intermediate budget level at which his expected payoff is maximized.

Within the context of a single firm, "When is the separation of operations and finance a good enough approximation of reality, and when should these functions be coordinated because the interactions are too important to ignore?" are the fundamental questions that Qiaohai (Joice) Hu, Lode Li, and Matthew Sobel address in Chapter 12, Production/Inventory Management and Capital Structure. They consider a setting where bankruptcy is costly, and hence the Modigliani-Miller Theorem no longer holds. Specifically, they study the integration of operational and financial decisions for a public firm that maximizes the expected present value of the dividends issued each period. In each period, the firm makes short-term decisions, including inventory (operational) and loan and dividend (financial), as well as long-term decisions, including long-term debt level. In a multiperiod framework, the chapter analyzes the optimal level of operational and financial decisions, the value of integrating the operational and financial decisions, and the relationship between the short-term and long-term decisions.

When the retailer does not have access to a perfectly competitive banking industry, the interaction between the retailer and the bank becomes strategic, with the bank turning into a profit maximizing lender. This is the scenario considered in Chapter 13, Bank Financing of Newsvendor Inventory: Coordinating Loan Schedules, by Qiaohai (Joice) Hu and Maqbool Dada. A key managerial insight is that the bank's profit tends to increase when the retailer's demand distribution becomes less skewed, and a nonlinear profit splitting loan mechanism can coordinate the financial supply chain of the bank and the retailer.

Operational Risk Management Strategies

Part IV, Operational Risk Management Strategies, consists of three chapters. In Chapter 14, Decentralized Supply Chain Risk Management, Göker Aydın, Volodymyr Babich, Damian Beil, and Zhibin Yang provide a taxonomy of supply risks and discuss challenges to manage those risks in decentralized supply chains where self-interested firms are interacting. Four key challenges are emphasized in managing supply risk in decentralized supply chains: misalignment of incentives between buyers and suppliers, competition among suppliers, competition among buyers, and asymmetric information. Through an in-depth review of literature they discuss the tools and trade-offs involved in the use of various operational risk management tools.

Focusing on the effectiveness of using supplier portfolios to cope with demand uncertainty, Victor Martínez-de-Albeníz studies the optimal sourcing decisions and investigates the trade-off between cost and flexibility in procurement in Chapter 15, Using Supplier Portfolios to Manage Demand Risk. A buyer contracts with two suppliers, a low-cost supplier offers a contract with no adjustment flexibility, and a high-cost supplier offers the flexibility that allows the buyer to adjust quantities after receiving demand forecast updates. The key managerial insight is that the firm should use the low-cost contract to fulfill the demand that materializes with high likelihood, and use the high-cost, but more flexible, contract to fulfill the fraction of demand that has higher uncertainty. The author also investigates the value of using portfolio of contracts over using a single contract and the impact of demand uncertainty on this value.

Inventory buffering is another effective tool to cope with demand uncertainty. In Chapter 16, An Opportunity Cost View of Basestock Optimality for the Warehouse Problem, Nicola Secomandi shows an easy-to-understand derivation of the optimality of the two-level basestock policy, a common trading practice in the commodity industries. The author illustrates that the results based on a simple modeling setting provides insights on the policies that should be adopted in more complex environments, the development of computational algorithms, and financial hedging policies.

Industrial Applications

Part V, Industrial Applications consists of four chapters. In Chapter 17, Procurement Risk Management in Beef Supply Chains, Onur Boyabatlı, Paul Kleindorfer, and Stephen Koontz study the commodity price risk management of a meatpacker in beef supply chains. In particular, they consider the optimal mix of contract and spot purchases in providing input (fed cattle) to a meatpacker from upstream feedlots and spot markets, when the meatpacker acts as a wholesaler into beef-product markets. They describe the background of the U.S. beef industry and provide computational results based on data for the U.S. beef industry described in the GIPSA Report (2007), and complemented the analysis by industry demand and supply studies. The analysis is focused on determining the impact on the optimal procurement portfolio, the expected profit, the value of spot and contract market and the expected plant utilization of spot price and demand uncertainty, the degree of substitution between products in final markets, as well as the cost characteristics of the meatpacker and the nature of quality and cost differences in the contract and spot markets. The chapter provides a foundation for understanding the complementary roles of contract and spot markets in U.S. beef markets.

In Chapter 18, Risk Management in Electric Utilities, Stein-Erik Fleten, Jussi Keppo, and Erkka Näsäkkälä review two main risks in electricity markets: volume risk and price risk. The authors provide a discussion on how to model and hedge these risks. They introduce a continuous stochastic process to model both types of risk. The optimal timing of a single trade of forward contracts is determined to minimize the variance of the cash-flow at the end of the planning horizon. Other practical risks in electricity markets, including liquidity risk and operational and political risk, are also discussed.

The final two chapters provide a holistic view of supply chain risk management from the perspective of industry practitioners. In Chapter 19, Supply Chain Risk Management: A Perspective from Practice, Colin Kessinger and Joe McMorrow offer a practical approach for assessing risk and resiliency in the supply chain: what is supply chain risk management (SCRM), how to develop SCRM capabilities, and what is SCRM process approach, and so on. They illustrate a case study of Cisco Systems' response to the 2008 Sichuan Earthquake.

Focusing on how to enable quick response after the occurrence of risk events, in Chapter 20, A Bayesian Framework for Supply Chain Risk Management Using Business Process Standards, Changhe Yuan, Feng Cheng, Henry Dao, Markus Ettl, Grace Lin, and Karthik Sourirajan develop a Bayesian graphical model to identify, quantify, mitigate, and respond to the risks affecting global supply chains. Based on the risk categorization network, which maps risk factors to business processes, this Bayesian model enables automatic learning using information such as business process standards, heterogeneous operational data, and expert knowledge. This methodology is further illustrated using a case study based on global logistics process performance data.

Acknowledgments

First and foremost, we sincerely thank all the authors who contributed to this handbook: Jan Van Mieghem, Vinod Singhal (Kevin Hendricks), Brian Tomlin (Yimin Wang), Melvyn Sim (Karthik Natarajan, Chung-PiawTeo), Qing Ding, Sridhar Seshadri (Çağri Haksöz), Bardia Kamrad, Mustafa Pınar (Alper Şen, A. Gökay Erön), Wenhui Zhao, René Caldentey (Xiangfeng Chen), Lode Li (Qiaohai (Joice) Hu, Matthew Sobel), Maqbool Dada, Volodymyr Babich (Göker Aydın, Damian Beil, Zhibin Yang), Victor Martínez-de-Albeníz, Nicola Secomandi, Paul R. Kleindorfer (Stephen Koontz), Jussi Keppo (Stein-Erik Fleten, Erkka Näsäkkälä), Colin Kessinger (Joe McMorrow), Grace Lin (Changhe Yuan, Feng Cheng, Henry Dao, Markus Ettl, Karthik Sourirajan). Their enthusiastic support to this project has made the process a great experience for us.

We want to thank our doctoral students at Olin Business School, Washington University:Ehsan Bolandifar, Xingxing Chen, Junghee Lee, Zhili Tian, Mike Mingcheng Wei, Xiaole Wu, Yixuan Xiao, Yueshan Yu, who read the earlier draft of the manuscripts and offered many insightful suggestions for improvement.

We are indebted to Paul R. Kleindorfer for encouraging and supporting this book project, and generously agreeing to write a foreword to this book.

We are grateful to the organizers and participants of the Mini-conferences on Integrated Risk Management in Operations and Global Supply Chain Management held at Washington University in St. Louis in 2004, 2005, 2007 (sponsored by The Boeing Center on Technology, Information and Manufacturing), at the University of Michigan in 2006, and at Lee Kong Chian School of Business, Singapore Management University in 2008. Those conferences promoted the early research in integrated risk management and helped shape this emerging field of research.

Contributors

GÖKER AYDIN, Kelley School of Business, Indiana University, Bloomington, Indiana, USA

VOLODYMYR BABICH, McDonough School of Business, Georgetown University, Washington, DC, USA

DAMIAN BEIL, Ross School of Business, University of Michigan, Ann Arbor, Michigan, USA

ONUR BOYABATLI, Lee Kong Chian School of Business, Singapore Management University, Singapore

RENÉ CALDENTEY, Stern School of Business, New York University, New York, New York, USA

XIANGFENG CHEN, School of Management, Fudan University, Shanghai, People's Republic of China

FENG CHENG, Federal Aviation Administration, Washington, DC, USA

MAQBOOL DADA, John Hopkins Carey Business School, Baltimore, Maryland, USA

HENRY DAO, IBM Global Business Services, Waltham, Massachusetts, USA

QING DING, Singapore Management University, Singapore

LINGXIU DONG, Olin Business School, Washington University, St. Louis, Missouri, USA

A. GÖKAY ERÖN, Department of Industrial Engineering, Bilkent University, Bilkent, Ankara, Turkey

MARKUS ETTL, IBM T.J. Watson Research Center, Yorktown Heights, New York, USA

STEIN-ERIK FLETEN, Department of Industrial Economics and Technology Management, Norwegian University of Science and Technology, Trondheim, Norway

ÇAĞRI HAKSÖZ, Faculty of Management, Sabancı University, İstanbul, Turkey

KEVIN B. HENDRICKS, School of Business & Economics, Wilfrid Laurier University, Waterloo, Ontario, Canada

QIAOHAI (JOICE) HU, Krannert School of Management, Purdue University, West Lafayette, Indiana, USA

BARDIA KAMRAD, McDonough School of Business, Georgetown University, Washington, DC, USA

JUSSI KEPPO, Department of Industrial and Operations Engineering, University of Michigan, Ann Arbor, Michigan, USA

COLIN KESSINGER, Managing Director, End-to-End Analytics, Palo Alto, California, USA

PAUL R. KLEINDORFER, Paul Dubrule Professor of Sustainable Development, INSEAD, Fontainebleau, France

STEPHEN R. KOONTZ, Department of Agricultural and Resource Economics, Colorado State University, Fort Collins, Colorado, USA

PANOS KOUVELIS, Olin Business School, Washington University, St. Louis, Missouri, USA

LODE LI, Yale School of Management, New Haven, Connecticut, USA; and Cheung Kong Graduate School of Business, Beijing, People's Republic of China

RONG LI, Lee Kong Chian School of Business, Singapore Management University, Singapore

GRACE LIN, Department of IEOR, Columbia University, New York, New York, USA; and Green Value Net, Inc., Chappaqua, New York, USA

VICTOR MARTÍNEZ-DE-ALBÉNIZ, IESE Business School, University of Navarra, Barcelona, Spain

JOE MCMORROW, Senior Manager, Global Business Operations, Customer Value Chain Management, Cisco Systems, Inc., San Jose, California, USA

ERKKA NÄSÄKKÄLÄ, Systems Analysis Laboratory, Helsinki University of Technology, Hut, Finland

KARTHIK NATARAJAN, Department of Management Sciences, College of Business, City University of Hong Kong, Kowloon, Hong Kong

MUSTAFA Ç. PINAR, Department of Industrial Engineering, Bilkent University, Bilkent, Ankara, Turkey

NICOLA SECOMANDI, Tepper School of Business, Carnegie Mellon University, Pittsburgh, Pennsylvania, USA

ALPER ŞEN, Department of Industrial Engineering, Bilkent University, Bilkent, Ankara, Turkey

SRIDHAR SESHADRI, McCombs School of Business, The University of Texas, Austin, Texas, USA

MELVYN SIM, Department of Decision Sciences, NUS Business School, National University of Singapore, Singapore

VINOD R. SINGHAL, College of Management, Georgia Institute of Technology, Atlanta, Georgia, USA

MATTHEW J. SOBEL, Weatherhead School of Management, Case Western Reserve University, Cleveland, Ohio, USA

KARTHIK SOURIRAJAN, ZS Associates, Chicago, Illinois, USA

CHUNG-PIAW TEO, Department of Decision Sciences, NUS Business School, National University of Singapore, Singapore

BRIAN TOMLIN, Tuck School of Business, Dartmouth College, Hanover, New Hampshire, USA

JAN A. VAN MIEGHEM, Kellogg School of Management, Northwestern University, Evanston, Illinois, USA

YIMIN WANG, W.P. Carey School of Business, Arizona State University, Tempe, Arizona, USA

ZHIBIN YANG, Lundquist College of Business, University of Oregon, Eugene, Oregon, USA

CHANGHE YUAN, Department of Computer Science and Engineering, Mississippi State University, Mississippi State, Mississippi, USA

WENHUI ZHAO, Antai College of Economics and Management, Shanghai Jiao Tong University, Shanghai, People's Republic of China

PART ONE

Foundations and Overview

CHAPTER ONE

Integrated Risk Management: A Conceptual Framework with Research Overview and Applications in Practice

PANOS KOUVELIS, LINGXIU DONG, ONUR BOYABATLI, AND RONG LI

1.1 Introduction

The past three decades have witnessed unprecedented growth in global trade. Firms in one part of the world can enjoy the great opportunity to access the input and output markets, technologies, and capitals in other parts of the world. The economically well-connected world exposes firms to multiple and high magnitude risks. The set of the normal business risks is expanded as firms are now often facing unfamiliar and uncertain demand and supply markets, and unanticipated commodity price shocks and currency exchange rate fluctuations. The previously perceived unusual business risks, such as unexpected supplier bankruptcies in a turbulent global economy, and supply disruptions as a result of physical disasters

Handbook of Integrated Risk Management in Global Supply Chains, First Edition. Edited by Panos Kouvelis, Lingxiu Dong, Onur Boyabatli, and Rong Li.

and terrorist attacks, have been intensified and cannot be neglected any longer as the frequency of these events has been increasing, posing a significant threat to business continuity. Although the existence of risk is not news to the business world (i. e., companies have long been buying insurances to protect them against certain risks) the development of advanced risk management tools did not start until the advancement of economic theory and information technology less than three decades ago, and its progress varies by industry.

Challenges in practice have motivated growing interests among academicians, especially researchers in supply chain management, in developing economically sound and practically feasible risk measures, understanding the influence of firms' risk preferences on supply chain decisions, and building multifunctional tools for risk management and control.

In this chapter we present an action-based supply chain risk management framework that has emerged from industry practice and academic research. For practitioners, this conceptual framework can serve as a guideline to devise risk management strategies that suit their specific supply chain environments. For academicians, we map the research in the growing field of integrated risk management research within the context of this framework, and identify potentially fruitful directions for future research.

1.2 An Action-Based Framework for Supply Chain Risk Management

The framework of supply chain risk management proposes a two-stage action plan: a *planning* stage and an *execution* stage. In the planning stage, carefully thought-out plans and proactive actions should be put in place to ensure business continuity and to sustain profitability in the event of an undesirable scenario. The main actions include identifying the prospective supply chain risks, assessing the likelihood of risk occurrence and the severity of consequences, and devising risk mitigation plans and putting counter measures in place to avoid or reduce (if possible) the probability of risk events and to reduce the damages/disruptions to the supply chain. In the execution stage, firms should establish a risk scanning mechanism to detect signs of risk events, put in place a real-time risk response process that is ready to deploy recovery plans immediately, and have a measuring system to assess all relevant data and analyze the effectiveness of the scanning and response processes.

We will discuss in more detail the activities involved in the two stages in the sections that follow.

1.2.1 PLANNING STAGE

Three steps of actions should take place in the planning stage: risk identification, risk assessment, and risk mitigation planning.

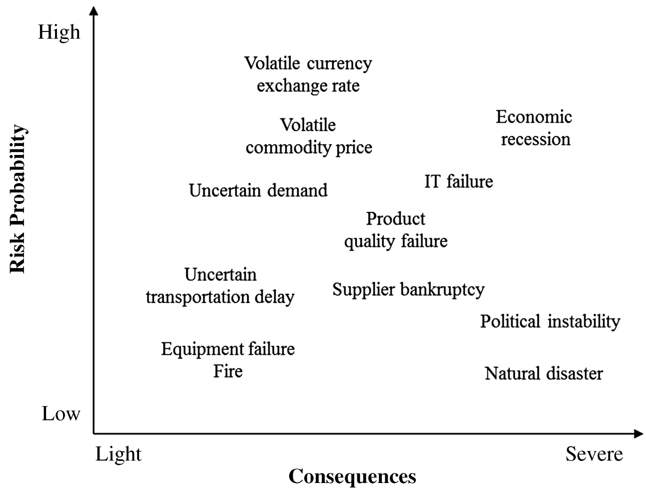


FIGURE 1.1 **Supply chain vulnerability map.**

1. **Risk identification.** This step identifies each possible adverse event, large or small, and produces a supply chain vulnerability map where the likelihood of those events and the severity of their consequences are estimated roughly. Risk identification is an important step of supply chain risk management and should involve multiple business functions and/or supply chain members, because this exercise not only raises the awareness of the various risks that the supply chain is exposed to, it also leads management to prioritize efforts in the following tasks of risk assessment and mitigation planning. Figure 1.1 provides an example of supply chain vulnerability map, where the vertical axis represents the probability that a risk event occurs and the horizontal axis represents the severity of the consequence.

   There are a number of ways to further analyze supply chain risks. One approach is to distinguish between the *internal risks* and the *external risks*. *Internal risks* are driven by the weakness of planning/control/coordination within the supply chain, for example, inaccurate demand forecasts, machine failures, uncertain yields, and supplier bankruptcy; *external risks* are driven by the events that are outside control of the supply chain members, for example, fluctuations in commodity prices, currency exchange and interest rates, natural hazards, war, and terrorism. Distinguishing between the internal and external risks directs firms to apply suitable risk avoidance and risk mitigation strategies to different types of risks.

   Another approach is to categorize the risks as *normal business risks* versus *disruption risks*. The *normal business risks* are driven by uncertainties inherent in the business with moderate-to-high frequency of occurrences, including uncertainties in supply and demand, and in costs and prices. The *disruption risks* are driven by relatively low-probability adverse events that can disrupt the normal function of the supply chain and lead to business discontinuity;

examples include natural or man-made disasters (earthquake, fire, flood, pandemic diseases, etc.) and other unforeseeable events. Accordingly, the goal of devising risk mitigation strategies will be different. Because of the reoccurring nature of the normal business risks, they are to be dealt with on the daily basis; a supply chain should be *robust* enough to perform effectively in a wide range of operating conditions. Disruptions, however, occur once in a while and organizations should develop the ability to quickly adapt to and overcome the disruptions; in other words, a supply chain should build in *resilience* to respond to the changing environments.

2. **Risk assessment**. Building upon the work of the risk identification step, this step quantitatively assesses the probability of the risk events and their implications for the supply chain. Several challenges can arise in performing this task:
   - Insufficient data for performing statistical analysis in order to derive the probability distribution of certain risk events. Expert opinions are often sought to assign such probabilities.
   - Risk events may be driven by some common factors, and therefore, their occurrences can be correlated. The process of identifying the underlying drivers and the likelihood they would lead to various risk events, and deriving the correlation among those risk events also often rely on experts' subjective inputs.
   - Risk events often have multifaceted impact on supply chain performance. Assessing the scope and the magnitude of their consequences requires, again, cross-functional, cross-organizational communication.

   The outcome of the assessment should help firms to prioritize risks—what risks are essential to control and mitigate, and what can be neglected.

3. **Risk mitigation planning:** Risk prevention plans and countermeasures are devised in this step to decrease the probability of adverse events and decrease the severity of the consequence.

   The discussion of the risk identification step suggests that using different categorization of supply chain risks allows firms to see the differences in the nature of the risk mitigation strategies. Understanding whether a risk is internal or external to the supply chain helps firms to set realistic expectations of what risk mitigation can achieve. For external risks, which are out of the firms' control, risk mitigation strategies should strive to reduce the magnitude of the negative impact, although in some circumstances, some external risks such as earthquake and flood can be avoided by careful design of the global facility network of the supply chain (e.g., by choosing facility locations far away from risk regions). For internal risks, whose likelihood is largely determined by the management of the supply chain, risk mitigation strategies should be more focused on identifying the root causes of those risks and deriving control or coordination mechanisms to prevent the risk events from happening.

   Viewing risks as normal business risks versus disruption risks, firms should be aware that risk mitigations plans are different in their relationship to the firms' overall supply chain management strategies. Risk mitigation strategies managing normal business risks are themselves an integral part of the supply chain strategy, and should be applied seamlessly in daily operations to respond to the anticipated, frequent, normal risk events. For disruption risks, because of the high uncertainty in when, where, what will happen, and the scope of the impact, the risk mitigation strategies cannot be designed precisely at the tactical level, because many pieces of the plan will be designed and executed 'òn the fly" in real-time as updated information of the specific disruption or crisis becomes available. The three processes of the execution stage (to be discussed below) are critical for such disruption management, although they are also relevant to managing normal business risks.

Risk mitigation strategy has been the main focus of the supply chain risk management research. Section 1.3 of this chapter will discuss in more detail the landscape of the current academic research in this area.

1.2.2 EXECUTION STAGE

The execution stage consists of three processes: scanning, response, and measurement.

1. **Scanning.** This process tracks what is happening in the supply chain in real-time and informs the appropriate executives immediately when risk events occur. Risks identified in the planning stage should be assigned to corresponding business functions and/or supply chain members to monitor; exception conditions should be defined to distinguish between normal business risks and disruptions, and communication infrastructure should be established to allow for instantaneous attention from management.
2. **Response.** Clear roles and responsibilities and the organization's flexibility to adapt to changing environments are two elements of the successful execution of this process. This process calls upon specific action plans when a normal risk event happens. For disruption risks, disruption management teams are assembled to design and execute recovery plans.
3. **Measurement.** This process documents and collects data throughout the organization to assess the effectiveness of activities and processes in the planning stage and the execution stage. For supply chain risk management to be effective, it must be treated as a part of the business process, constantly being revisited and improved upon. Thus, the measurement process closes the loop of the supply chain risk management in two important ways. First, the data documented can be used to facilitate the analysis in risk identification and assessment; second, the analysis conducted after each risk mitigation event allows the organization to learn, to find the strength and weakness in its risk management ability, and to shed light on areas for improvement.

1.3 Risk Mitigation Strategies

The growing research of supply chain risk management can be largely divided into three themes: managing risk through operational strategies, integrated operations-finance risk management, and supply chain finance. We briefly discuss the philosophy and approaches in each research theme.

1.3.1 MANAGING RISK THROUGH OPERATIONS STRATEGIES

Many risk events in supply chain cause delays and disruptions in matching supply with demand. This area of research is mainly focused on operations strategies that ensure the availability of resource (inventory and capacities) to cope with variations in supply and demand conditions. A nice framework to view those operational strategies in four categories, based on the timing when the resources are becoming available at firm's disposal, can be found in Hopp (2007). Table 1.1 provides the definition and examples of each category.

As the costs of carrying out those strategies are different, situations in which they are most effective are also different. Figure 1.2 maps the operational strategies with situations identified by the likelihood of risk event and the severity of consequences (Hopp 2007).

When the severity of consequences is light, it is not worthwhile to adopt any risk mitigating strategies. Buffering and pooling strategies require the firm to keep the physical resources ready for use when a risk event occurs. Uncertain supplier lead times require the firm to plan the orders with excessively large order lead times. When the likelihood of risk events is low or moderate, those strategies are not economically justifiable. Hence, ideally, buffering and pooling strategies should be adopted when risk events are highly likely and their consequences are not too light.

For situations where the likelihood of risk events is moderate, contingency planning (securing the access to resource in times of need through advanced planning or contracts) is a more reasonable strategy to adopt, provided that the

TABLE 1.1 Four Categories of Operational Risk Mitigation Strategies

	Definition	Examples
Buffering	Maintaining excess resources	Safety stock, idling capacity, safety leadtimes
Pooling	Sharing of resources	Flexible technology, common components, transshipment, postponement
Contingency planning	"Virtual buffering" established by a preset course of action	Backup supply, multisourcing/supplier portfolio

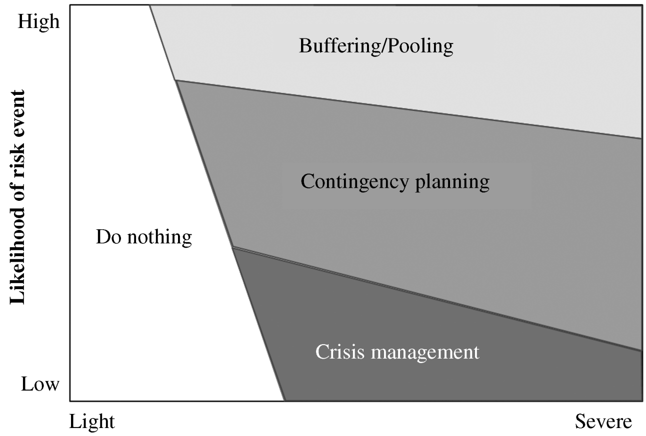


FIGURE 1.2 **Matching risk mitigation strategies to types of supply chain risk (Hopp 2007).**

cost of doing so is lower than buffering or pooling. Having access to a reliable and responsive backup supplier or buying from a spot market are typical examples of such plans.

For situations where risk events are rare but consequences are severe, crisis management will be needed when resources from buffering/pooling and contingency planning are insufficient to maintain the normal functioning of the business.

Buffering, pooling, and contingency planning have received a great deal of attention in the academic research, with quantitative models deriving the optimal execution of each strategy and comparing effectiveness of various strategies. Research on crisis management remains scarce in supply chain management research, because crisis management, by definition, requires the real-time planning and execution of recovery for unanticipated events. Building organizations capable of such responses has attracted some attention in the organizational behavior and design literature, but often not within a supply chain disruption context.

1.3.2 INTEGRATED OPERATIONS-FINANCE RISK MANAGEMENT

This theme of research recognizes the diverse set of risks that firms face today, and some of those risks, such as fluctuations in commodity prices and currency exchange rates can be mitigated using financial derivative contracts. This research argues that an integrated risk management approach via coordinated operational and financial decisions offers an overall optimal risk control strategy, and provides insights on the important questions of executing these functionally aligned strategies: How to achieve the coordination of these decisions? What are the value drivers of an integrated risk management? Are operational hedging and financial hedging strategic substitutes or complements?

Since risk aversion is one of the main drivers for firms' financial hedging activities, this line of research often incorporates different risk-averse utility functions, such as mean-variance, value-at-risk, and exponential, to explore the trade-offs of the expected payoff and the spread of undesirable outcomes implied by operational and financial hedging strategies.

1.3.3 SUPPLY CHAIN FINANCE

Most of the existing supply chain management research assumes that firms' operational decisions and financial decisions are made independently, and focus on the coordination of the material flow and information flow across supply chain members, leaving the question of how material flows are financed outside the scope of the research. This operations-finance independence assumption is justified by the well-known Modigliani–Miller Theorem, which states that under a set of assumptions a firm's value is unaffected by how the firm is financed. Those assumptions include rational investors, the absence of transaction costs, corporate taxes, and asymmetric information, and thus, a perfect capital market. In practice, those assumptions seldom hold simultaneously, and supply chains decisions more often are affected by how they are financed, and by payment terms, and the pressure of improving cash flow efficiencies. The supply chain finance research explicitly considers the constraints imposed by limited capital, lack of credit for borrowing, taxation, asymmetric information, and so on, and explores the trade-offs involved in making operational and financial decisions, and proposes schemes to coordinate the material, information, and financial flows within the supply chain.

1.4 Research Opportunities

While so far most of the academic research in supply chain risk management is concentrated in understanding and developing risk mitigation strategies, there are still many opportunities for more interdisciplinary research in this area. In integrated risk management, the interface research should develop tailored models to institutional details involved in development of financial hedging strategies, and provide insights on the classification of important uncertainties to hedge, and on the main operational and financial drivers of a profitable integrated risk management program based on industry characteristics.

More research is needed in understanding the value of coordinating financial and operational decisions. One particular domain of interest is the procurement and risk hedging of multitude of commodities. There exist empirical observations about firms centralizing their financial risk management functions over multiple commodities, and at the same time, delegating the physical sourcing of each commodity to separate divisions (i.e., in a decentralized fashion). It would be interesting to analyze conditions where it is better to centralize/decentralize operational and financial decisions, and the implications on the firm's overall performance.

In supply chain finance, opportunities exist in understanding the relationship between the supply chain inventory policy/ownership and the management of the cash-to-cash cycle, their impact on firms' financial performance, and how the answers are different to firms at different stages of business life cycle, and/or with different risk attitude. As more innovative supply chain financing schemes are being offered by financial institutions, a framework for mapping business environments with the most effective financing schemes would be of greater value.

In both of the above research areas we are only scratching the surface of the large body of finance, economic, and accounting literature. We see great potential in collaborating with colleagues in those areas to develop a more integrated approach to tackle those challenges, as also should be done in practice.

In the operations management field, several papers analyze different operational hedging strategies of firms and delineate their value in specific problem domains (such value of flexible technology in multiproduct firms, value of production switching among different subsidiaries, etc.). The insights generated from these papers tend to depend on the particular setting under consideration. A future challenge is to develop a robust classification of operational hedging strategies with their structural properties relevant for a multitude of operational settings. The general methodology and insights developed in the finance literature for financial hedging is a potential starting point for this purpose. On the other side of the picture, the finance literature can make use of the knowledge base on operational hedging developed in the operations management literature. One immediate future research direction is developing empirical proxies for the measurement of operational hedging capabilities of firms. In the finance literature, the level of operational hedging is generally attributed to the dispersion, that is, the number of different locations of the subsidiaries of the firms. A higher dispersion implies better operationally hedged firms. However, as demonstrated by several scholars in the operations literature, a less dispersed firm can be better operationally hedged than a more dispersed firm if the internal operational flexibility (such as flexible production technologies) of the former is higher.

Academic research in risk assessment and real-time crisis management remain scarce. While academic research has heavily relied on the presence of historical data to assess probability and severity of consequences, the nature of many supply chain risks makes the availability of such relevant data a scarce resource. Climate changes are making even weather events hard to effectively estimate from historical data, and the unpredictability of terrorist plots reduces the importance of previous observations.

Finding ways to estimate probabilities from events that happened in the past, which are not exactly parallel to the ones we are trying to predict using pattern recognition approaches that reveal similarities beyond the immediate obvious; using experts that can bring different domains of knowledge to understand a future event; and exploring the interdependency between our mitigation actions of some future risks and their effects on the probabilities and consequence levels of these events are important directions of future investigation.

Although there has been a great deal of research on risk measures and utility theory in economics, mathematics, and finance literature, there is lack of research on appropriate and consistent risk measures and utility functions to characterize overall supply chain risks, combining operational and financial risks. When different risk-averse utility functions are adopted for a same problem, different results may be obtained. Comparative statistics should be provided for various risk measures and utility functions in the context of supply chain management. Empirical research can be conducted to help understand objectively how firms perceive, measure, and value each supply chain risk. Guidance should be provided to the practitioners on what utility functions are appropriate for what type of supply chain, where the type may be determined by, but not limited to, the magnitude of the risks, nature of risks, and interaction of these risks.

There is a need to place more emphasis on understanding the development of organizational capabilities and leadership in managing risk events and crisis situations. In environments of high uncertainty and unpredictable nature events, risks have to be managed in ways substantially different than our traditional project risk management approaches. There is the need to develop organizational units that are very flexible in their structures, which can identify and solve problems in rapid cycles, when communication and reporting happens very frequently and often informally. While our traditional risk planning and mitigation approaches emphasize redundancy of resources and portfolio approaches in risk mitigation, the organizational responses will rely on tightly linked systems that cherish interdependency and the risk of failure as motivators for outside-of-the-box thinking and fast, innovative responses using all available talents and resources leveraging the intimate understanding of individual player capabilities. Research that identifies organizational designs, training approaches, team building methodologies, incentive systems, and leader development pathways to effective crisis management is a vital missing block in our current integrated risk management theory. Actually, practice and effective corporate responses witnessed in certain recent crises can guide theory development in this area.

REFERENCE

Hopp W.J. 2007. Supply Chain Science. New York: McGraw-Hill Irwin.

CHAPTER TWO

Risk Management and Operational Hedging: An Overview*

JAN A. VAN MIEGHEM

2.1 Introduction

This chapter, which is based on Chapter 9 in Van Mieghem (2008), aims to give an introduction and overview of risk management and the techniques that operations managers can use to mitigate risks. We start the next section by describing the concept of risk management and viewing it as an ongoing four-step process and integral part of operations strategy. We distinguish operational from financial risk. In Section 2.3, we identify the various operational risks that companies are exposed to. We then review methodologies to assess and value those risks both qualitatively (using subjective risk maps) and quantitatively (using risk preference functions and risk metrics). The goal of risk assessment is to improve how we react to risk and to proactively reduce our exposure to it. In section 2.5, we review tactical risk decisions, including risk discovery and risk recovery. The remaining

*The author is grateful for the thoughtful and constructive suggestions by the the three anonymous referees.

Handbook of Integrated Risk Management in Global Supply Chains, First Edition. Edited by Panos Kouvelis, Lingxiu Dong, Onur Boyabatli, and Rong Li.

sections of the chapter illustrates strategic risk mitigation (i.e., how operations can be structured to mitigate specific risks).

Hedging refers to any action taken to mitigate a particular risk exposure; operational hedging uses operational instruments. Section 2.7 posits that there are four generic strategies to mitigate risk using operational instruments: (1) reserves and redundancy, (2) diversification and pooling, (3) risk sharing and transfer, and (4) reducing or eliminating root causes of risk. Section 2.8 reviews financial hedging of operational risk using options and derivatives. Section 2.9 illustrates how operational hedging can be tailored to the specific operations strategy of the firm using techniques such as: tailored redundancy, dynamic pooling with allocation flexibility, chaining, and multisourcing. Section 2.10 finishes the chapter by summarizing some guidelines for operational risk management.

2.2 Risk Management: Concept and Process

2.2.1 DEFINING HAZARDS AND RISK

Before we can describe the concept of risk management, we must first define some terms. *Hazards* are potential sources of danger. In a business setting, danger can mean anything that may have a negative impact on the firm's net present value. Hazards have a harmful impact, but they may or may not occur.

In everyday language, *risk* refers to an exposure to a chance of loss or damage. ("We risked losing a lot of money in this venture"; "Why risk your life?") Risk thus arises from hazards *and* exposure: it does not exist if exposure to a hazard does not or will not occur (e.g., if you live on top of a mountain, you are not at risk of flooding). The interpretation of risk as *an undesirable possible consequence of uncertainty* suggests that risk is a combination of two factors:

1. The probability that an adverse event or hazard will occur.
2. The consequences of the adverse event.

2.2.2 FINANCIAL VERSUS OPERATIONAL RISK

While it is intuitive to associate risk with a probability and an undesired outcome, there are other interpretations of risk. The 1997 Presidential-Congressional Commission on Risk Management defined risk as the probability that a substance or situation will produce harm under specified conditions. In economics, "risk refers to situations in which we can list all possible outcomes and we know the likelihood that each outcome occurs" (Pindyck and Rubinfeld 1989).

In finance, risk is "the possibility that the actual outcome is likely to diverge [or deviate] from the expected value" (Sharpe 1985). In finance, risk is equated with uncertainty in payoffs, which we will refer to as *profit variability risk*. Risk then implies the existence of some random variable whose standard deviation or variance can be used as a measure of risk. Notice that this view calls any uncertainty in outcomes, whether favorable or not, risk. The key distinction from the common interpretation of risk is the absence of "danger" or an "adverse event." For instance, people don't typically say that they are at risk of winning the lottery.

Operational risks are risks that stem from operations (i.e. from activities and resources). Any potential source that generates a negative impact on the flow of information, goods, and cash in our operations is an operational risk. The inclusion of cash flowing through the operation implies that financial and operational risks are not mutually exclusive. But the goal of operations is to maximize expected firm value by matching supply with demand. Any possible mismatch between supply and demand, excess or shortage, is undesirable and is called *mismatch risk*.

2.2.3 RISK MANAGEMENT: CONCEPT AND EXAMPLES

In general, *risk management* is the broad activity of planning and decision making designed to deal with the occurrence of hazards or risks. Risks include both unlikely but high-impact disruption risks, as well as more common volatility in demand, internal processing, and supply.

Procter & Gamble provides an example of managing disruption risk. On Sunday May 4, 2003, 1,200 workers at the company's Pringles plant in Jackson, Tennessee, heard warning sirens and rushed to evacuation areas. About 18 minutes later, tornados hit and badly damaged the plant's roof, while subsequent rain damaged truck loads of potato chips. The south end of the building was demolished and required reconstruction. With the sole Pringles plant in the Americas shut down, P&G had no choice but to suspend all U.S. distribution, armed with only a six-week supply of Pringles already in stores or en-route. It was estimated that it would take at least one month before shipments could resume, causing a huge blow to one of P&G's biggest brands. (According to the company, people eat 275 million chips per day, generating annual sales above $1 billion.) But the company was prepared: by 3 a.m., the brand contingency team and an entire recovery process (described in Example 2.1) was set in motion. We shall return to the importance of tactical risk management through fast risk discovery and recovery.

EXAMPLE 2.1 **Risk Management by Procter & Gamble**

Only hours after a tornado hit P&G's Pringles plant in Jackson, Tennessee on Sunday May 4, 2003, the brand contingency team started the recovery processes. Employees from the only other Pringles plant in Mechelen, just outside of Brussels, were flown in to help reconstruction. By Wednesday, P&G determined that its major equipment would be fine, and put its major

U.S. customers on allocation. By Saturday, a temporary roof had been installed; on Monday, May 12, a limited production of its most popular flavors was resumed.

Meanwhile, production in Belgium was maximized and re-routed to supply some of the Jackson plant's Latin American and Asian customers. According to the Mechelen plant:

> "Already in the second week of May, first Raw & Pack Material orders were placed at our suppliers with stretched leadtimes which enabled Mechelen to switch its production schedule by the end of the third week (the 2 lines with the capability to run Asian product—14 case count versus 18 case count—started to run the Asian brand codes).
>
> "First, shipments to the Asian market left Mechelen by the end of May! In total Mechelen delivered 11,100,000 200g cans and 7,500,000 50g cans! On top of this achievement, Mechelen produced specific flavors for Japan that were never ran before (a special Operations-QA-PD team was formed to qualify our lines for these specific flavors).
>
> "As a consequence of this massive support, the inventories in Mechelen for the Western European market were heavily eroded. Due to this low inventory the Mechelen organization was further stretched to provide good service levels for Western Europe. We discovered some opportunities in our supply chain (which would be more difficult to find when they were hidden under stock).
>
> "Net: Mechelen protected the Asian business with huge flexibility and strengthened its own supply chain by doing that" (Van Campenhout 2004).

Strategic risk mitigation involves the structuring of global networks with sufficient flexibility to mitigate the impact of hazards. For example, BMW enjoys demand risk mitigation through its global operations network by building cars in Germany, Britain, the U.S., and South Africa. Out of the annual 160,000 Z4 roadsters and X5 sport "activity" vehicles built in 2003 in its Spartanburg, South Carolina plant, about 100,000 were exported, mostly to Europe. At the same time, BMW imported about 217,000 cars from Europe to reach annual U.S. sales of about 277,000 cars.

Partial balancing of flows through global manufacturing networks such as those of BMW or DaimlerChrysler's service networks (e.g., large consulting and accounting companies) can also mitigate currency exchange risk. For example, Michelin, the world's biggest tire maker, drew 35% of its 2003 annual sales from North America. While this would normally expose the French company to dollar-euro currency exchange risk, Michelin was not worried about exchange rates. They compensated for the loss caused by translating American revenues into euros by purchasing raw materials that are priced in dollars.

In contrast, companies like Porsche which builds cars mostly in Germany, must raise local prices to make up for currency changes (a dangerous approach that almost wiped Porsche out in the U.S. in the early 1990s). Otherwise, it must absorb the changes in the form of lower profits, or may resort to financial hedging instruments that we will describe below.

2.2.4 RISK MANAGEMENT AS A PROCESS AND INTEGRAL PART OF OPERATIONS STRATEGY

Now that we know what is meant by risk, we can proceed with the topic of this chapter: managing risk through operations. It is useful to think of risk management as a four-step process, as illustrated in Figure 2.1:

1. **Identification of hazards.** The first step in any risk management program is to identify the key potential sources of risk in the operation.
2. **Risk assessment.** The second step is to assess the degree of risk associated with each hazard. Then we must prioritize hazards and summarize their total impact into an overall risk level of the operation.
3. **Tactical risk decisions.** This step describes the appropriate decisions to be taken when a hazard is likely to occur soon, or when it has already occurred. For high risk levels, these decisions are also called "crisis management."
4. **Implement strategic risk mitigation or hedging.** This step involves structuring the operational system to reduce future risk exposure.

To adapt to change and to incorporate learning and improvement, risk management must be approached as a process; these four steps must be executed and updated recurrently.

It is useful to make a distinction between tactical and strategic risk management. Tactical risk management uses mechanisms to detect whether a specific hazard is likely to occur soon; then, it executes contingency plans. For instance,

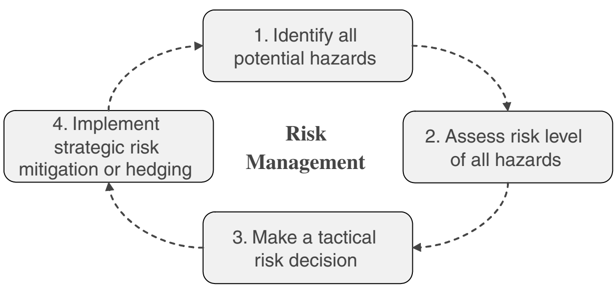


FIGURE 2.1 **Risk management as an ongoing process with four steps.**

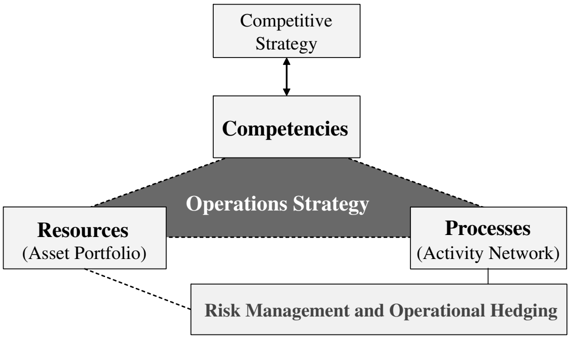


FIGURE 2.2 Operational hedging is a process of strategic risk mitigation. It involves structuring resources and processes to reduce future risk exposure. Therefore, operational hedging is an integral part of operations strategy.

P&G used warning sirens and followed a contingency plan to deal with the tornado strike on May 4, 2003, in Jackson, Tennessee (Example 2.1).

In contrast to dealing with the occurrence of a specific hazard, strategic risk management is concerned with mitigating future risk exposure. Operational hedging, a subset of strategic risk management, refers to the adjustment of strategies and the structuring of resources and processes to proactively reduce, if not eliminate, future risk exposure. For instance, P&G's Pringles operations comprise two manufacturing plants with sufficiently flexible processes enabling them to partially take over each other's work. This operational system provides a form of insurance that resulted in the tornado strike having limited financial impact.

In summary, operational hedging is an integral part of operations strategy (Fig. 2.2) for two reasons: It is a necessary process in each operation, and it involves structuring the entire operational system. The remainder of this chapter will illustrate the four-step process of risk management, meanwhile describing how risk management interacts with the operational system's resources and other processes.

2.3 Identification of Operational Hazards

The first step in any risk management program is to identify any potential sources of danger. According to one manager who participated in many risk assessment processes: "One lesson I learned is that hazard identification is one of the most difficult steps in the process. Without a clear and robust framework, it is nearly impossible to identify all critical hazards." Now, we will describe one approach to help this identification process.

An organization is most affected when a danger affects its ability to serve the customer's needs. Although a danger might impact an operation, the effects on the organization and its future are limited if the customer does not suffer from that impact. To identify important risks, it is useful to adopt the customer's perspective by asking: What is my customer's worst nightmare?

The answer then can be linked to operational risks that stem from our activities and assets. As described Van Mieghem (2008), any operation can be viewed from three perspectives: as a bundle of competencies, processes, or resources. Adopting these three views directly suggests three approaches that should be combined to identify operational hazards.

2.3.1 IDENTIFYING OPERATIONAL RISKS USING THE COMPETENCY VIEW

Linking competency failures to customer nightmares is probably the most direct way to focus the mind on important operational risks. What is the impact of a failure in the firm's key competencies such as quality, flexibility, timeliness, cost, or quantity? If operations strategy is well aligned, this importance should correspond to the priority ranking of the competencies in the customer value proposition. While this link is direct, it is not directly actionable. Therefore, the competency risks must be linked to processes or resources so we can restructure processes and resources to mitigate the competency risks.

2.3.2 IDENTIFYING OPERATIONAL RISKS USING THE PROCESS VIEW

Potential hazards can be identified and categorized by considering each activity in the value chain, as shown in Figure 2.3. Depending on the stage in the value chain where the negative impact may happen, we have:

1. *Innovation risk* represents any exposure to hazards that originate during research and development. The pharmaceutical industry provides a good example: a new drug or compound may turn out to not have sufficient efficacy, potency, or safety to be approved by the relevant governmental agency. Another example is Intel, which recently pulled the plug on the development of a 3Ghz Pentium chip, its fastest microprocessor for personal computers, because it proved to be too difficult to manufacture.
2. *Commercial risk* represents any exposure to hazards that originate in marketing and sales and negatively impacts revenues. It includes the risk that new products or services are not adopted, cash risks (e.g., lower sales prices than expected), or receivables risks (when customers don't pay).

FIGURE 2.3 Identifying operational risks using the value chain.

3. Closely related are *demand and supply risks*, which refer to any uncertainty in quantities demanded or supplied for a given product or service at a given time. Typical examples include retail risks, in which case we may have leftover stock that must be discounted, or insufficient supply (stockouts, underages). Supply risks may also refer to *sourcing risk*, which stems from interaction with suppliers. It may include risks in information (the wrong order was communicated or the order was not received), risks in goods (the wrong quantity or quality of goods was received), or risks in cash (the supply ends up being more expensive than expected). For example, a supplier may claim not to have received an order, or may have sent the wrong amount or type of supply. The shipment may have been lost or stolen. A supplier may have a capacity or yield problem, or may even undergo a catastrophic event such as terrorism, sabotage, or financial bankruptcy.
4. *Production and distribution risks* include any exposure to hazards that originate in our internal processing and distribution networks. There may be labor issues, worker safety hazards and non-ergonomically designed work environments, or maintenance failures that affect capacity availability. Inventory may be at risk of spoilage, damage, or loss. Unexpected operator errors, yield problems, accidental damage, and delays may increase cost above expectations. Distribution channels may be at risk of logistics provider failure, route or transportation mode disruptions, and other hazards (similar to sourcing risks).
5. *Service risk* refers to the exposure to hazards during after-sale service interactions. This may include lack of procedures to deal with product returns, problems, and service inquiries.
6. *Coordination and information risks* refer to uncertainty in coordination and information. They may stem from internal miscommunication and often result in internal demand-supply mismatches. Examples include information technology system failures in hardware, software, local, and wide area networks. Other information risks include forecasting risks, computer virus risks, and errors during order-taking and receiving.

Some industries, such as the pharmaceutical industry, also use the term *technical risk* to refer to the innovation risk of launching a new technology or drug. It is distinct from ongoing operational risk and commercial risk: while a drug may be approved and be no longer at technical risk, it still remains to be seen whether it will have sufficient demand at reasonable prices for it.

2.3.3 IDENTIFYING OPERATIONAL RISKS USING THE RESOURCE VIEW

One can also consider each asset in the operational system and identify associated potential hazards. In practice, one would investigate the key assets in the operation. We can classify assets, and corresponding risks, into three types:

1. *Capital asset risks* are exposures to hazards originating from property, plants, and equipment. These include exposures to property and environmental liability, equipment unreliability, as well as financial risks related to maintenance and perhaps future resale. They can also include working capital such as inventory and receivables risk.
2. *People risks* include safety, health, operational dependence, operator and management errors, resignations, turnover, absenteeism, sabotage, stealing, and more.
3. *Intangible asset risks* include policy risks, intellectual property risks, reputation, culture, and more.

2.3.4 SURROUNDING BACKGROUND RISKS

No organization operates in a vacuum. Aside from operational risk, the operating system is subject to various hazards that originate from its surroundings. Depending on the source, we can categorize types of background risks as:

1. *Natural risk*: In addition to operation-specific hazards, nature is capricious and can expose organizations to natural hazards such as earthquakes, heavy rains, lightning, hail storms, fires, and tornados. The exposure typically depends on the location of the organization. For example, coastal properties are exposed to coastal storm hazards such as hurricane storm surges, flooding, erosion, and wind.
2. *Political risk*: This risk includes any negative, unexpected change in laws and regulations (political stability is typically preferred). Examples include a breach in business contracts without recourse to legal action, unexpected strengthening in environmental or labor laws, unexpected currency devaluations, or an outbreak of war.
3. *Competitive and strategic risk* refers to the potential negative impact of competitors' actions, or environmental and technological changes that reduce the effectiveness of the company's strategy.

2.3.5 WHO SHOULD IDENTIFY POTENTIAL HAZARDS?

Everyone involved in the operation should be able to identify potential hazards. Naturally, people closest to the activities or assets often have the best knowledge. For example, account managers, service representatives, and technicians are most knowledgeable in identifying service risk. In contrast, supplier relationship and purchasing managers are the natural parties to identify sourcing risk. This means that risk identification requires a multifunctional team that can interact with functional specialists.

2.4 Risk Assessment and Valuation

The second step in any risk management program is to analyze the degree of risk associated with each hazard. The goal of risk assessment is to indicate which areas and activities in the value chain are most susceptible to hazards.

2.4.1 QUALITATIVE RISK ASSESSMENT: THE THEORY

Recall that risk is an undesirable consequence of uncertainty. Risk assessment thus involves, for each hazard identified in step 1, the estimation of:

1. The impact (vulnerability) on the organization if the hazard were to occur.
2. The probability of the hazard occurring during the operation.

The result can be displayed in a *subjective risk map*, an example of which is shown in Figure 2.4. The word "subjective" reminds us that this risk map is based on expert opinion only and not on statistical analysis. Obviously, the risk map is company specific: The risks carry different weights depending on the competitive strategy and the industry. For example, for a commercial bank, IT systems failure would have a much greater impact than would a hurricane.

Risk assessment is completed by ranking hazards to locate the highest-risk activities. This can be done qualitatively by combining the impact and probability

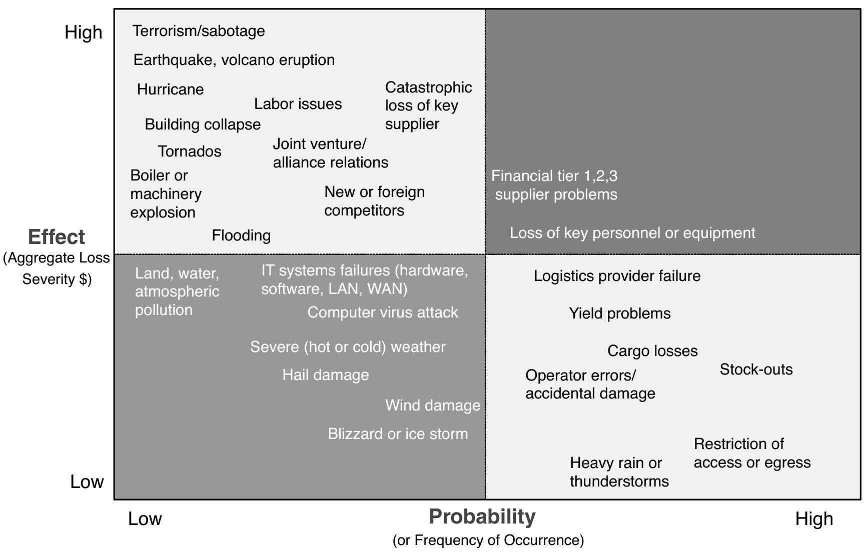


FIGURE 2.4 A subjective risk map is a graphical representation of the risk assessment for a specific organization done with the help of expert opinions. It shows the impact versus the likelihood of occurrence for each hazard.

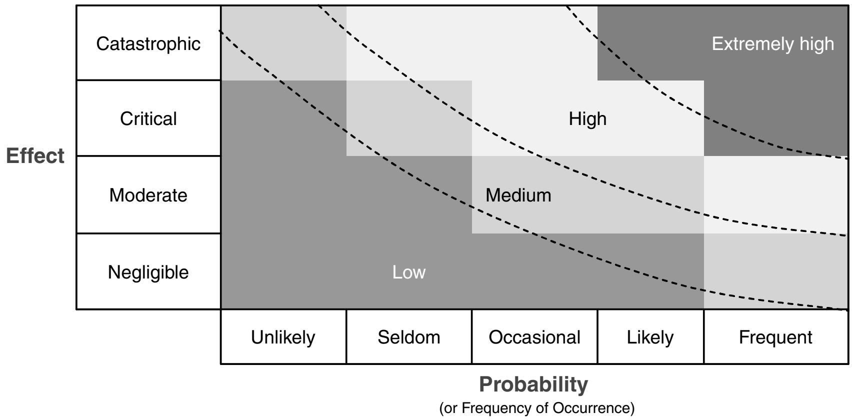


FIGURE 2.5 Qualitative risk assessment assigns an overall risk level to each hazard, depending on its probability and its impact.

for each hazard into an overall risk level. The risk map in Figure 2.4 classifies hazards into three risk levels. High risk hazards occupy the upper right quadrant and create high damage with a high probability. Medium risks are unlikely hazards with high impact (also called *disruptions*) or frequent, low impact hazards (recurrent risks). Low risks stem from unlikely hazards with low impact, and occupy the lower left quadrant.

2.4.2 QUALITATIVE RISK ASSESSMENT: EXAMPLES FROM PRACTICE

Smart operations managers periodically assess risks. For example, Figure 2.5 shows how the National Interagency Fire Center (2002) assigns risk levels (extremely high, high, medium, or low) for helicopter operations depending on the hazard's probability or frequency (unlikely, seldom, occasional, likely, or frequent) and impact (negligible, moderate, critical, or catastrophic).

Debit card companies and other financial companies conduct risk assessment programs periodically. According to one debit card product manager:

> "We had to go through *every* possible operational risk to our business annually, provide an estimate of impact of a hazardous event (on a scale of 1 to 5, covering a range of dollar values) as well as the likelihood of the event happening (also on a scale of 1–5). If you provided a top-two score high impact and high probability event, you were asked to present to the bank's risk management committee, which consisted of senior and executive managers and was headed by the bank's newly formed enterprise risk manager. They would expect to see your action plans if the event occurred, as well as the steps you've taken to mitigate the risk."

> "As part of the BASEL II requirements, all banks must conduct this type of thorough assessment for all areas of their business. Failure to meet the BASEL standards can result in sanctions by banking oversight committees (Fed, OCC, etc.) that could affect a bank's abilities to lend, to lend at good rates, to get approval for M&A, etc. It is quite an exhaustive accounting of operational risks. Admittedly, many estimates were just educated guesses by line managers and, of course, it also took a lot of time out of managers' days to focus on events that most likely weren't going to happen... In the end, though, the risk assessment process helped everyone realize where we were vulnerable. It also helps bank management have a much broader understanding of the entire risk exposure and brought operational risk management to the executive board level."

Some risks, such as political risks, are difficult to assess, compared to calculating the technical risk of product approval or the statistical risk of poor forecasting. Yet, where there is a will, there is a way. According to one risk assessment team, "one way to help dimension political risk is to compare the political risks of one country relatively to the risks faced in other countries the firm operates in. One team member found research that provided political risk indexes for various countries throughout the world. Other resources to help quantify what seemed to be a rather nebulous topic include the World Bank's Multilateral Investment Guarantee Agency and numerous consulting firms and insurance providers."

2.4.3 QUANTITATIVE RISK ASSESSMENT: RISK METRICS

The qualitative approach can be quantified by estimating the financial impact and probability of each hazard from past data and experience. A hazard's "risk level" can then be quantified by its expected impact, which is equal to the financial impact multiplied by the probability of occurrence. Constant risk levels are then represented by hyperbolic curves in risk maps, as illustrated by the dotted lines in Figure 2.5.

Besides the methods that assess the expected value of a hazard, there are many other ways of quantifying risk. These are most easily described by letting X denote the (financial) effect of a hazard or random event (i.e., X is a random variable) and $\overline{X}$ its mean or expected value $\mathbb{E}X$. Recall that financially, risk is considered to be the possibility that actual outcomes deviate from expected ones. A basic risk metric is *variance* (or its square root, the standard deviation), the expected squared deviation around the mean:

$$\text{variance} = \mathbb{E}(\overline{X} - X)^2 = \sigma^2$$

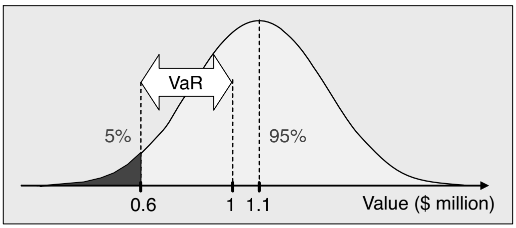
Variance and standard deviation treat positive deviations from the mean ("the upside") symmetrically with negative deviations ("the downside"). Statistical measures that exclude upside deviations are arguably more natural metrics of risk, because they only capture the undesirable consequences of uncertainty. A popular downside risk measure is *Value-at-Risk* (VaR). It measures the worst expected loss at a given confidence level by answering the question: How much can I lose with $x\%$ probability over a preset horizon? Example 2.2 illustrates how to calculate VaR. Other examples of downside risk metrics are:

$$\text{below-mean semivariance} = \mathbb{E}((\overline{X} - X)^{+})^{2}$$
$$\text{below-target } t \text{ semivariance} = \mathbb{E}((t - X)^{+})^{2}$$
$$\text{expected below-target } t \text{ risk} = \mathbb{E}\,(t - X)^{+}$$

where the notation X^{+} means the positive part of X, that is, $X^{+} = \max(0, X)$.

EXAMPLE 2.2 How to Calculate Value-at-Risk (VaR)

Value-at-Risk at $x\%$ is the answer to the question: how much can be lost with $x\%$ probability over a preset horizon? Suppose you currently have a portfolio worth \$1 million, and its annual return is normally distributed with mean 10% and standard deviation 30%. What is your value-at-risk at 5%?



Calculating value-at-risk at 5%.

Your value-at-risk at 5% can be calculated in two steps, as illustrated in the figure above:

1. Find the 5% quantile of next year's value. In our example, Excel gives us that number as `norminv(.05,1.1,.3)`= \$0.6 million.
2. Find the VaR as the difference between the 5% quantile of future value and the current value. In our example, the VaR is $1 - 0.6 =$ \$.4 million.

This means that there is only a 5% chance that you will lose more than \$400,000.

2.4.4 VALUING RISK WITH PREFERENCES AND UTILITY FUNCTIONS

Measuring risk directly in terms of the downside volatility of outcomes is certainly informative, but such raw risk metrics do not allow us to easily compare risks. For example, do you prefer a risky project with a value variance of \$1 to another with a variance of \$100? Surely, you would want to know the expected value before answering! As a matter of fact, if your preferences depend on expected values *only*, you are said to be *risk-neutral*.

Most people, however, are *risk-sensitive*, which means that their preferences do not depend only on expected value. Deciding between two risky projects then requires trading off risk with expected return. Making this trade-off is difficult in general, but under standard rationality assumptions we can use a utility function to summarize risk preferences. A utility function u simply maps outcomes into a decision-maker's utility. A risky outcome X_1 then is preferred over outcome X_2 if and only if the expected utility of the first exceeds that of the second.

It directly follows that a risk-neutral manager would have a linear utility function, so that only expected outcomes matter. For example, consider choosing between two projects: the first project has a payoff of \$100 for sure, while the second's payoff has an expected value of \$100, but is normally distributed around that mean with standard deviation σ. A risk-neutral manager derives equal expected utility from both projects and is indifferent between them.

In contrast, risk-averse managers have concave utility functions, which reflect their higher sensitivity to downside than upside. To see this, consider a concave function such as the negative exponential

$$u(x) = 1 - e^{-\gamma x}$$

shown in Figure 2.6. The parameter $\gamma > 0$ represents the manager's sensitivity to risk and is called the *coefficient of absolute risk aversion*. As the coefficient of risk aversion γ increases, the utility function becomes more concave and more sensitive to downside variations. Notice that the upside has a maximal utility of

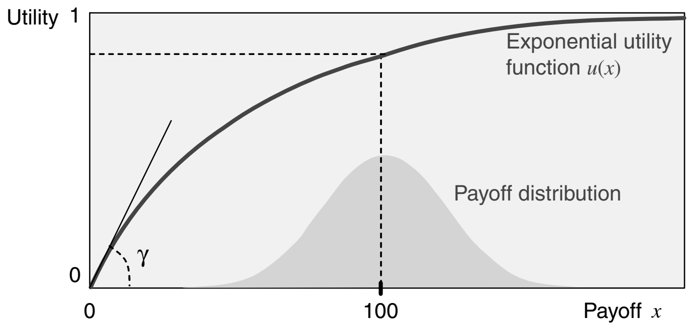


FIGURE 2.6 **Risk-averse managers have concave utility functions, thus preferring a unit reduction of the downside to a unit increase of the upside.**

1, while the downside is unlimited. The marginal utility of \$1 above the mean is less than that of \$1 below the mean. In other words, a risk-averse manager gets *more* utility from reducing the downside by one unit than from increasing the upside by one unit. It follows that downside variation is not offset by equal upside variation, and that the expected utility from a random outcome with mean 100 is strictly less than a certain outcome of 100. A risk-averse manager dislikes volatility.

2.4.5 MEAN-VARIANCE FRONTIERS

Risk-averse valuation with expected utilities typically requires calculus, but there is one useful exception. When payoffs are normally distributed, their expected exponential utility can be expressed by the simpler *mean-variance preference*:

$$\text{mean-variance preference } MV = \mu - \frac{\gamma}{2}\sigma^2$$

where μ is the expected payoff and σ^2 is its variance. Expected utility increases with the mean payoff, but decreases if the actual outcome is more likely to deviate from its expected value (as indicated by a greater variance) or if the manager is more risk-averse (as indicated by a greater coefficient of risk-aversion).

Mean-variance preferences are at the core of modern financial *portfolio management* and provide a good inspiration for operations strategies for risk mitigation. The original idea was first formulated in 1952 by Nobel laureate Harry Markowitz, who employed mean-variance preferences. He started by observing that individual investors are not interested in the expected value of their portfolio only. If that were the case, portfolios would consist of one asset only: that with the highest expected return.

Most investors hold diversified portfolios because they are concerned with risk as well as expected value. Markowitz used the variance of portfolio value as a measure of risk. Not only are mean-variance preferences reasonable models to describe the decisions of a risk-averse investor, but variances of a portfolio are also easily computed as a function of the covariances between any pair of assets in the portfolio. Markowitz thus presented a mathematical approach to optimal portfolio selection depending on the investor's risk-aversion, represented by the coefficient of risk aversion γ.

Optimal portfolio selection can be illustrated graphically as follows. Imagine that you calculated the expected value (return) and variance (risk) of all possible portfolios that can be bought with a given budget. Now represent each portfolio by one point on a risk-return graph, as shown in Figure 2.7. Then, the optimal portfolio can be derived in two steps:

1. Only portfolios that lie on the northwestern frontier, called the *mean-variance frontier*, should be selected; these are called *efficient* portfolios (any other portfolio is dominated by an efficient one with the same expected return but less risk, or the reverse).

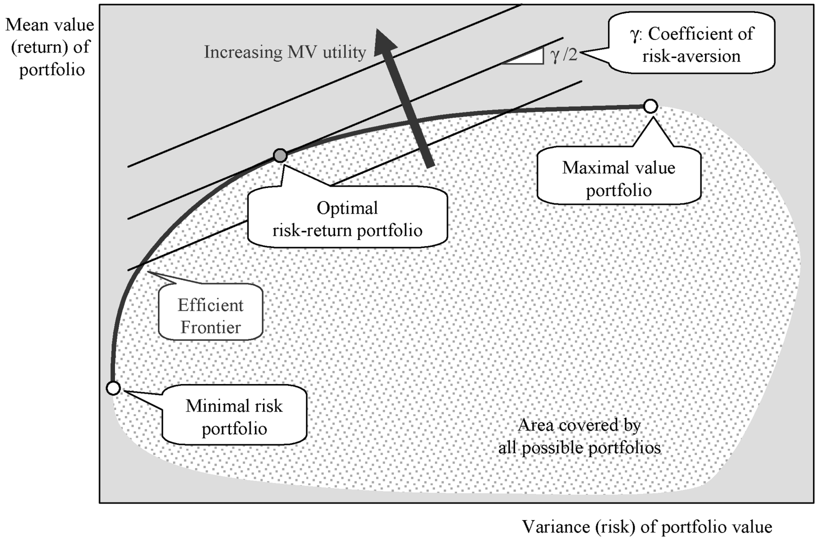


FIGURE 2.7 Graphical representation of Markowitz's optimal portfolio selection. The optimal risk-return trade-off for a manager with coefficient of risk aversion γ is the point on the efficient frontier with a tangent of $\gamma/2$.

2. Once the frontier is known, the final step is to estimate the investor's coefficient of risk-aversion γ, in order to identify the optimal portfolio with the point on the frontier with tangent $\gamma/2$. Indeed, the investor maximizes his expected utility by maximizing his mean-value preferences, which are straight lines in the risk-return graph.

Before we use this approach for strategic risk mitigation and operational hedging, we must consider step 3 of the risk management process.

2.5 Tactical Risk Decisions and Crisis Management

The third step of the risk management process is to let our risk assessment guide us in developing a plan of appropriate tactical decisions to be taken when a specific hazard is likely to occur, or when it has already occurred. For high risk levels, these decisions are also called "crisis management." Tactical risk management involves three activities: risk preparation, risk discovery, and risk recovery.

2.5.1 RISK PREPARATION

To be successful in later risk recovery, organizational risk preparedness is key. This means that companies have formulated proactive and reactive plans: what to do if risk levels are elevated, and contingent actions to take after the hazard occurs.

In addition to making plans, they also practice the plans with fire drills, backup routines for power losses, and so on.

2.5.2 RISK DISCOVERY

In order to execute proactive plans, one must monitor risks and have a fast system of hazard detection or discovery. Reconsider the P&G Pringles plant example (see Example 2.1. When it became likely that the plant was in the path of an oncoming tornado, management decided that the risk level was sufficiently high to evacuate the plant. The anticipatory risk decision was to turn on the sirens as a signal to everyone that the earlier-designed evacuation procedure was in effect.

EXAMPLE 2.3 Risk Discovery and Recovery: Nokia v. Ericsson

At 8 pm on Friday, March 17th, 2000, a lightning bolt hit an electric line in New Mexico and, somehow, resulted in a fire at the Philips NV's semiconductor plant in Albuquerque. While the sprinkler system extinguished the fire in less than 10 minutes, it also destroyed the clean room during that process, and with it, millions of cell phone chips that were destined for its two largest customers, Nokia and Ericsson. But how the two companies responded to the crisis couldn't have been more different.

At Nokia, computer screens indicated delays of shipments from some Philips chips even before Philips called Nokia's chief component-purchasing manager Tapio Markki on Monday, March 20. Philips said the fire impacted some 4 million handsets and that there would be a one week delay. Given that it was about to introduce a new generation of cell phones based on the Philips chips, Nokia decided to further look into the issue and offered to fly two Nokia engineers to Albuquerque to help with the recovery. Philips declined the offer and said on March 31, two weeks after the fire, that they would need more weeks to repair the plant, and that several months worth of chip supplies could be disrupted.

Nokia went into textbook crisis management mode. Of the five parts, two were indispensable: one was made by various suppliers around the globe, while the other one was an applications specific integrated circuit (ASIC) made only by Philips. A Nokia team, headed by current chairman Ollila flew to Philips' headquarters in Amsterdam and spoke directly with Philips' CEO, Cor Boonstra in an attempt to find alternate supply. Nokia demanded capacity information about all Philips plants and insisted on rerouting the capacity. "The goal was simple: For a little period of time, Philips and Nokia would operate as one company regarding these components," said Nokia's Korhonen. As a solution, Philips used its plants in Eindhoven to produce more than 10 million units of the ASIC chip, and also freed up a Shanghai plant for Nokia. Meanwhile, Nokia engineers redesigned some of their chips so they could be produced elsewhere, and they worked further with Albuquerque to boost production.

Ericsson, in contrast, treated the initial call from Philips as "one technician talking to another." When Ericsson's top management finally learned about the problem several weeks later, it was too late. Philips had no more spare capacity left and no other suppliers were capable of providing the parts Ericsson needed. Thus, Ericsson came up millions of chips short in a rapidly moving cell phone market. The company said they lost at least $400 million in potential revenue. At the end of 2000, its mobile phone division announced a staggering $1.7 billion loss and vowed that it would never be exposed like this again. In January 2001, Ericsson exited the handset production business completely. Source: Latour 2001.

2.5.3 RISK RECOVERY

Once disaster has struck, risk recovery executes contingency actions such as finding other suppliers, temporarily changing prices to ease demand, providing substitutes when actual demand significantly differs from plan, or using backup suppliers or processes. For example, when Grainger, which supplies maintenance and operating parts, had its East Cost facilities hit by electricity blackouts or hurricanes in Florida, they switched to internal power generators; by using this quick backup strategy, Grainger did not miss a single order fulfillment. Similarly, once a tornado struck at P&G, managers immediately started a recovery operation by calling the corporate brand contingency team.

Fast risk discovery and recovery is paramount to containing the negative impact of a disruption. The differential reaction to the unforeseen problems at a Philips semiconductor plant by two of its customers, Nokia and Ericsson, provides a case in point (Example 2.3). Nokia quickly switched sourcing to other backup facilities and suppliers with little impact to ongoing operations, while Ericsson's slow response along with its unhedged single sourcing strategy is reported to have cost it $400 million in lost sales.

In summary, in good tactical risk management, companies are prepared, use risk discovery mechanisms, and have quick risk recovery plans (Figure 2.8). By the very nature of a crisis, however, there still is a fair amount of unforeseen decision making to be done. The first step is to examine options for addressing the risks. Then, make decisions about which options to implement. Finally, take actions to implement the decisions. Naturally, the appropriate decision maker for these contingent decisions is more senior the higher the risk level.

2.6 Strategic Risk Mitigation

Fast risk discovery and recovery from actual disruptions is paramount in containing negative impact. The effectiveness of such tactical risk decisions to respond

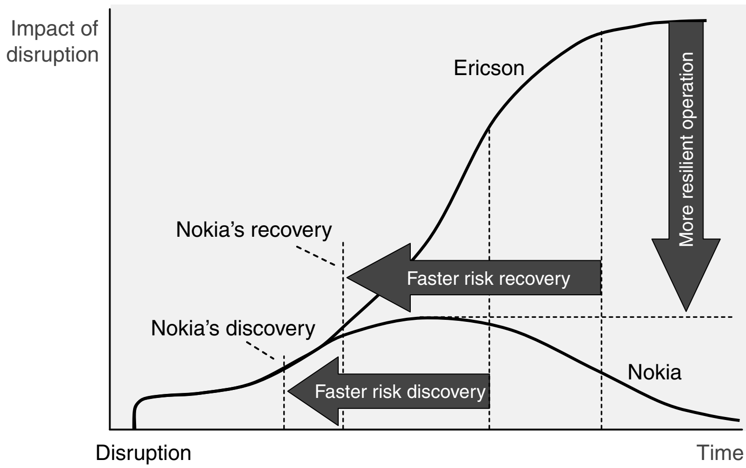


FIGURE 2.8 **Faster risk discovery and recovery, along with a more resilient operation, is paramount in mitigating the impact of disruptions (adapted from Sheffi 2005).**

to actual disruptions also greatly depends on the flexibility of the operational system. Crisis management is similar to operating a hospital's emergency room: Speed and flexibility are the most important competencies to quickly deal with unforeseen problems. Strategic risk management, the fourth step in risk management, involves configuring the operational system for speed and flexibility so as to mitigate future risk exposure. Its goal is to design what Sheffi (2005) calls a resilient organization.

Typically, it costs money to mitigate risk exposure. Strategic risk mitigation must balance that cost with the benefits of reduced risk exposure. The greatest benefit is typically gained by focusing on the most risky hazards (that were identified in step 2 of the risk management process) first. Let us discuss how to carry out the cost-benefit analysis behind strategic risk mitigation.

2.6.1 THE VALUE-MAXIMIZING LEVEL OF RISK MITIGATION (RISK-NEUTRAL)

Risk mitigation strategies fall on a continuum between risk acceptance and risk elimination. Many hazards have such a small risk that one simply accepts their exposure. For example, passengers and freight forwarders accept the inherent risks of flying. Sometimes, risks can be eliminated. For instance, P&G could eliminate tornado risk by relocating its plants to areas where tornados are highly improbable.

Typically, the marginal benefit of risk mitigation decreases while its cost increases, so that the appropriate risk mitigation level falls in between the extreme strategies of risk elimination and risk acceptance. For example, consider mitigating shortage risk by adding safety stock. Figure 2.9 depicts the costs and benefits of reducing shortage risk by adding safety stock for product with a sales price of \$5, a unit cost of \$1, and a normally distributed demand forecast with mean

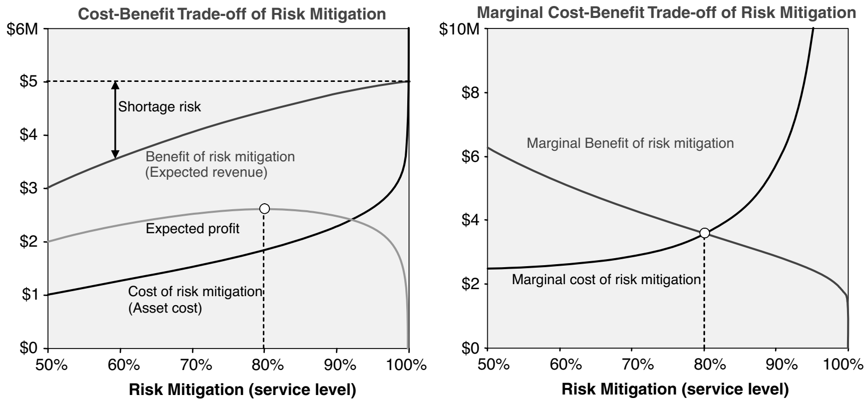


FIGURE 2.9 Risk-neutral managers determine the optimal level of risk mitigation by trading off costs and benefits. Consider adding safety stock in order to mitigate shortage risk. Initially, the increase in expected revenue outweighs the increase in inventory cost, which in turn increases expected profit (left panel). But beyond a certain level, marginal costs exceed marginal benefits (right panel).

and standard deviation of 1 million. When stocking the average demand, the shortage probability is 50%. Shortage risk mitigation requires adding safety stock. Complete shortage risk elimination would yield expected revenues of $5 million, but would require exorbitant safety stock. A risk-neutral manager is better off mitigating 80% shortage risk because that maximizes expected profits, according to the newsvendor model.

2.6.2 STRATEGIC RISK-RETURN TRADE-OFFS FOR RISK-AVERSE MANAGERS

Risk-averse managers care about profit risk as well as expected profits. They are willing to give up some expected profits for a reduction in profit risk.[1]

When managing a single asset such as capacity or stock, profit risk can be decreased by reducing the asset level. Reconsider our earlier example of mitigating shortage risk by adding safety stock. With an abundance of stock, shortages are eliminated and sales equal demand. The manager then is exposed to total demand risk, and profit standard deviation is maximized (equal to $5 × 1 million demand standard deviation). By reducing the stocking (and thus service) level, sales are capped by inventory and profit risk decreases (to $5 × the standard deviation of the minimum of demand and stocking level). Figure 2.10 depicts the mean and variance of profits as a function of the service level. Using Markowitz's approach, the appropriate level of risk mitigation for a manager with a coefficient of risk

[1] This is simply a statement of fact, not a prescription. In fact, managers of publicly held companies should maximize expected value, because shareholders can diversify risk on their own by engaging in portfolio management consistent with their own risk-reward preferences.

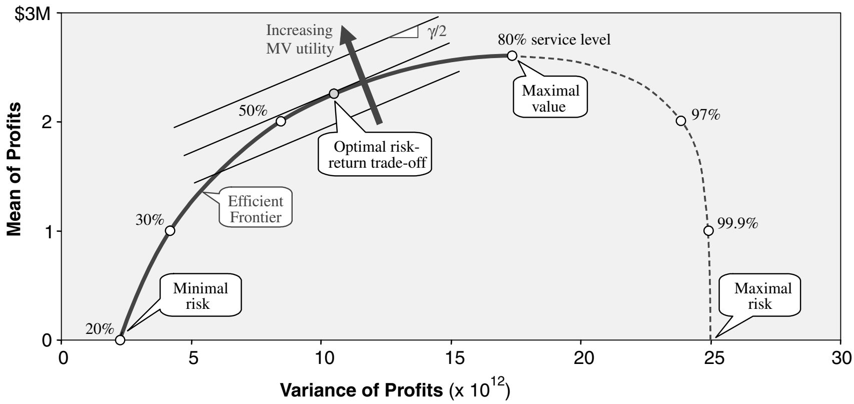


FIGURE 2.10 **Increasing inventory or capacity increases service level and reduces shortage risk, but increases profit variance risk. The optimal risk-return trade-off for a manager with a coefficient of risk aversion γ is the point on the efficient frontier with a tangent of $\gamma/2$.**

aversion γ is specified by the point on the frontier with a tangent of $\gamma/2$. By moving southwest along the frontier, we give up some expected profit and thus decrease profit variance risk.

2.6.3 PERIODIC UPDATING AND CONTINUOUS RISK MANAGEMENT

Strategic risk mitigation includes a procedure to keep risk assessment up to date. Business risks continually change over time and risk management must evolve accordingly. Just like periodic financial portfolio rebalancing and health checkups, periodic updating of risk management is a smart preventative move. Grainger, for instance, reviews each risk plan every six months and updates if there is a business change. It also performs real tests, as well as "desktop exercises" of its risk plans on an annual basis.

In the remainder of this chapter, we will turn our attention to various strategies that can be used to mitigate operational risk.

2.7 Four Operational Hedging Strategies

Hedging refers to any action taken to mitigate a particular risk exposure. It often involves counterbalancing acts that take on one risk to offset another. Most businesses hedge or insure to reduce risk, not to make money. In theory, a perfect hedge eliminates risk without impacting mean value. In practice, however, hedging impacts both risk and value. Using Markowitz's visualization, hedging becomes more effective as the frontier becomes flatter, so that risk reduction only comes with a small value loss.

The insurance industry uses three means to mitigate its risk: it builds reserves to meet claims, pools risks over many clients (this diversification reduces its total risk), and transfers remaining risks to reinsurers using contracts. Operations can also use these three generic risk mitigation strategies; in addition, there are also an arsenal of operations management techniques to reduce risk.

These four generic strategies to mitigate risk using operational instruments, (i.e. *operational hedging*), are summarized in Example 2.4. Let us review these four strategies qualitatively; the remainder of this chapter will quantify and tailor them to a particular situation.

EXAMPLE 2.4 Four Generic Operational Hedging Strategies

1. **Reserves & Redundancy**
   - safety capacity, safety inventory, safety time, warranties (reserves)
   - multisourcing, multiple locations and transportation modes, backup assets and processes (redundancy)
2. **Diversification & Pooling**
   - operating in diverse markets (diversification)
   - serving diverse markets with one resource (demand pooling)
   - using diverse suppliers for one resource (supply pooling)
   - allocation flexibility of suppliers, designs, resources, activities, and outputs
3. **Risk-Sharing & Transfer**
   - alliances and partnerships
   - outsourcing with structured supply contracts
   - entering financial hedging contracts with third parties
4. **Reducing or Eliminating Root Causes of Risk**
   - postponement with quick response (decrease risk exposure)
   - supplier collaboration and improvement
   - root cause analysis and variance reduction (Six Sigma, total quality management)
   - robust product and process design, including process relocation

2.7.1 RESERVES AND REDUNDANCY

A core risk mitigation strategy is to invest in reserves, which are assets held in excess of expected requirements, “just-in-case.” Reserves are well-understood and a key tactic in operations management: standard inventory and queueing models directly specify how risk-neutral decision makers should size safety capacity, safety inventory, and safety (lead) time as a buffer against uncertainty.

In general, redundancy refers to an excess over normal requirements or duplication. In engineering, redundancy is the duplication of critical system component to increase system reliability, often through backup assets or processes, such as Grainger's backup power generators. In the normal course of operations, these redundant assets or processes are not needed.

2.7.2 DIVERSIFICATION AND POOLING

Diversification refers to serving multiple risks (e.g., product demands) from one portfolio or network. This popular risk mitigation strategy is also known as "not holding all eggs in one basket." There are several ways of pooling risks with operations, each with a different impact:

1. *Pure diversification and natural hedging* refers to serving two markets with separate, dedicated resources. This reduces total profit variance risk because variability in one market partially offsets variability in the other (unless both risks are perfectly positively correlated). Supplying countries from local operations is an example of pure diversification that is also known as *natural hedging*. It mitigates profit variance risk arising from local demand risk as well as currency exchange risk. Notice that pure diversification does not impact expected value and differs from reserves and redundancy.
2. *Demand pooling* refers to serving multiple demands from one resource, such as a centralized warehouse that stocks one product to serve multiple areas, or a single facility that supplies multiple markets. Similarly, *supply pooling* means serving one demand from multiple suppliers; a typical example is multisourcing of a single component, Chopra and Sodhi (2004) and Tomlin B. (2006) provide several approaches that involve demand and supply pooling.

   Demand and supply pooling are special forms of diversification and risk-pooling. By "betting on two horses," they provide the profit variance risk mitigation benefits of pure diversification that are valued by risk-averse investors. In addition, they reduce expected mismatch costs, safety capacity, and safety inventory, while improving service (because resource sizing is driven by the aggregate standard deviation.) Thus, in contrast to pure diversification, demand and supply pooling also brings benefits to risk-neutral managers (but less so as correlation increases or if risks have dissimilar magnitudes).
3. *Allocation flexibility and information updating* refers to pooling *heterogeneous* risks with a *flexible* network. The embedded real options achieve more powerful operational hedging than do static demand or supply pooling. For example, consider serving continental Europe and the United Kingdom from a single process in Belgium. If the process has sufficient flexibility to postpone country allocations, it can first observe actual demands and exchange rates and then maximize revenues by steering the allocation to the more profitable country. It is exactly this type of dynamic pooling that was effective in the Pringles and Nokia examples.

   In addition to the embedded risk mitigation of pooling, allocation flexibility and other real options can increase expected profits. called this the revenue maximization option of flexibility, which becomes more valuable as the pooled risks become more heterogeneous. This "active" operational hedging highlights an interesting advantage over financial hedging or pure insurance: Operational hedging not only mitigates risk but can also add value by exploiting upside variations. We will illustrate this quantitatively in the next few sections.

Redundancy and diversity through flexible networks are related. For example, consider P&G's network for Pringles production has two plants. Each plant's main mission is to serve its own geography, so that neither plant is redundant, strictly speaking. The flexibility embodied in the network, however, does allow the Belgian plant to serve as a backup for the Jackson plant, illustrating its relationship to redundancy.

2.7.3 RISK SHARING AND TRANSFER

Instead of bearing all the risk ourselves, we can share it with partners, alliances, or suppliers. A vast supply chain contracting literature studies various structured contracts (e.g., buyback and revenue sharing contracts) that balance risk between a supplier and buyer. Later in this chapter, we will discuss how a company can share and even transfer risk by entering into financial hedging contracts with third parties. The obvious example of sharing risk is taking on insurance contracts.

2.7.4 REDUCING OR ELIMINATING ROOT CAUSES OF RISK

In addition to these three insurance-like techniques, operations research has also emphasized risk reduction by quick response, supply chain collaboration and continuous improvement. Continuous improvement uses root cause analysis and an entire arsenal of techniques for variance reduction. While reviewing those techniques go beyond the scope of this chapter, it cannot be overemphasized that, in the long run, eliminating problems is better than mitigating their impact. The Toyota Production System exactly tries to achieve this.

2.8 Financial Hedging of Operational Risk

Financial hedging uses financial instruments to mitigate risk. Let us discuss some examples of how financial hedging can mitigate operational risk and how it relates to operational hedging.

2.8.1 HEDGING DEMAND RISK WITH OPTIONS

Demand for discretionary items such as apparel, consumer electronics, and home furnishings is often correlated with economic indicators. Gaur and Seshadri (2005) present evidence that the correlation can be quite significant. For example, The Redbook Average (a seasonally adjusted average of same-store sales growth in a sample of 60 large U.S. general merchandise retailers representing

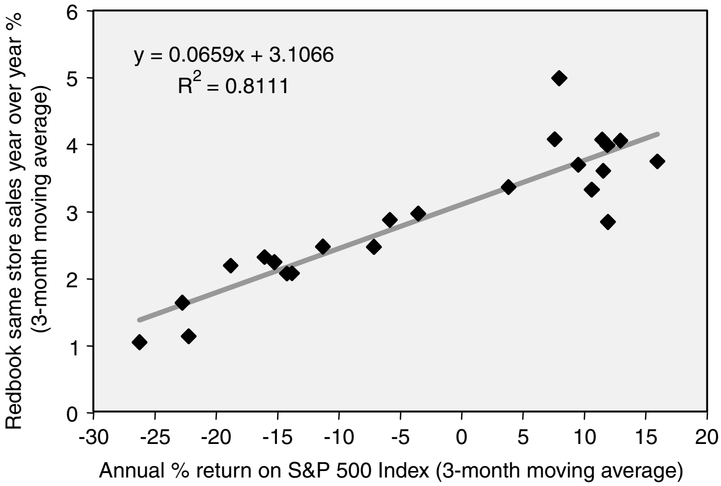


FIGURE 2.11 **Redbook's same-store sales growth rate is highly correlated with the annual return on the S&P 500 Index (Gaur and Seshadri 2005).**

about 9,000 stores) monthly time-series data from November 1999 to November 2001 had a correlation coefficient of 0.90 with same-period returns on the S&P 500 index ($R^2 = 81\%$, see Figure 2.11). In addition, that value of R^2 is correlated with the fraction of discretionary items sold as a percentage of total sales.

Similar results hold on the firm-level. Figure 2.12 shows that sales per customer transaction and sales per square foot at The Home Depot (a retail chain selling home construction and furnishing products) are both significantly correlated with the S&P 500 index.

Theoretically, the correlation between sales and a financial instrument can be exploited to mitigate demand risk by buying a (tailored) call option on the

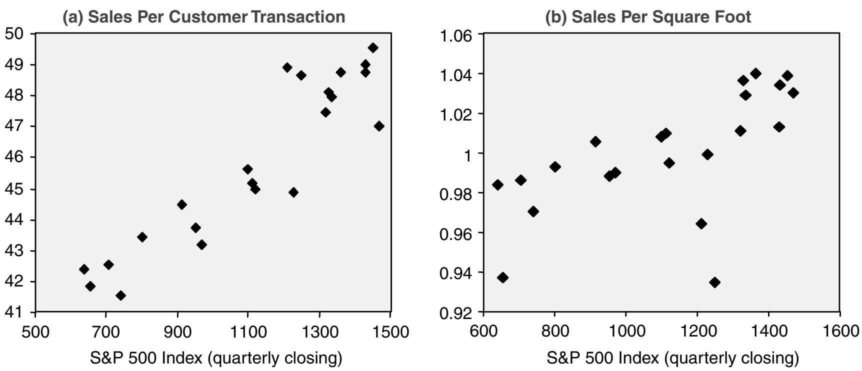


FIGURE 2.12 **Quarterly sales per customer transaction (a) and per square foot (b) at Home Depot are correlated with S&P 500 Index (Gaur and Seshadri 2005)**

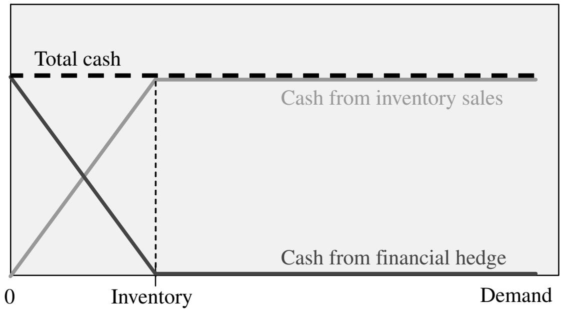


FIGURE 2.13 **When demand is correlated with a financial asset, inventory risk can be mitigated by buying a call option on the financial asset. If the correlation is perfect, a financial option with an exercise price corresponding to the inventory would provide a perfect hedge.**

financial asset.[2] Consider a retailer who must order inventory today but faces a leadtime of 4 weeks. Assume for simplicity that demand is perfectly correlated with the S&P 500 index. Buying call options on the index with exercise price corresponding to the inventory and exercise date one month from now would provide a perfect hedge, as shown by Figure 2.13.

In reality, the correlation is imperfect and the hedging transactions are more complex (involving a tailored family of different calls) but we can take away the main insights: financial hedging can significantly mitigate profit variability risk. (For this specific example, however, a healthy dose of caution is appropriate, given that Home Depot is part of the S&P 500; thus, the correlation is to be expected and may not be a reliable predictor of future performance.)

In well-functioning financial markets, arbitrage arguments show that the options are priced at a level equal to their expected return. Financial hedging then reduces variance risk without impacting the expected return. Risk-averse retailers will then increase their order sizes closer to the risk-neutral (newsvendor) level. In addition, financial market information can be used to update demand forecasts.

2.8.2 HEDGING DEMAND RISK WITH (WEATHER) DERIVATIVES

When demand is correlated with weather conditions, demand risk can be mitigated using financial weather derivatives. For example, Japanese insurer Mitsui Sumitomo sells derivatives based on the snowfall in a particular location. Retail ski shops could use that derivative to hedge against low snowfall that could impact sales. At the same time, Mitsui Sumitomo could sell the opposite derivative to a snow removal company. This example shows how intermediaries such as Mitsui can sometimes improve markets by balancing risks.

[2] A call option gives its owner the right to buy the asset at a specified exercise price on or before the specified exercise date.

In addition to snowfall, weather derivatives can include specifications on rainfall, temperature, and wind. In 2002, Mitsui Sumitomo issued a weather-derivative contract to a soft drink wholesaler based on the number of hours of sunshine. If the number of sunshine hours recorded in the July–September quarter fell below a certain predetermined threshold, Mitsui would pay the company a predetermined amount.

2.8.3 HEDGING CURRENCY RISK WITH FORWARD CONTRACTS AND SWAPS

Global firms like Michelin, BMW, and Porsche are exposed to currency exchange rate risk. Two popular risk mitigation strategies are natural hedging (produce and sell locally) and financial hedging involving forwards and swaps.

In a *foreign exchange forward market*, you can buy and sell currency for future delivery. If you are going to receive €500,000 next month, you can insure yourself by entering into a one-month *forward contract*. The forward rate on this contract is the price you agree to receive in one month when you deliver the €500,000. Forward contracts specify future customized bilateral transactions. (They can be used to hedge various types of risk. For example, Dong and Liu (2007) studied the equilibrium forward contract on a nonstorable commodity between two firms that have mean-variance preferences over their risky profits and negotiate the forward contract through a Nash bargaining process.)

More generally, you can manage risk by entering into a *swap*, which is a contract between two parties specifying the exchange of a series of payments at specified intervals over a specified period of time.

For main currencies and specific amounts and delivery dates' rates, there are standardized contracts, called *futures*, that are traded on currency future markets. In well-functioning financial markets, arbitrage arguments imply that future rates equal the expected rate so that forwards and futures do not impact expected value (neglecting small transaction costs). Inspired by an example of Professor John R. Birge, Example 2.5 illustrates how a global manufacturer can benefit from natural, operational, and financial hedging. As predicted, natural and financial hedging reduce profit variance without affecting expected profits. In contrast, active operational hedging can use allocation flexibility in the global network to produce and sell at the most advantageous location, thereby increases expected profits (combining that with financial hedging further reduces variance without impacting expected value).

EXAMPLE 2.5 Should We Use Financial Hedging with Futures, or Operational Hedging?

> Consider a global manufacturer with production facilities in Europe and the U.S. that is exposed to demand and currency risk. The firm wonders whether it should hedge financially or operationally.

Suppose that the unit sales price is €20,000 in Europe and $20,000 in the U.S.; similarly, the unit cost is 10k in local currency. Suppose that currencies are correlated with demand and that there are two states of nature, each equally likely:

1. U.S. demand is 100k units, Euro demand is 50k, and the exchange rate is $1/€.
2. U.S. demand is 50k units, Euro demand is 100k, and the exchange rate is $2/€.

Hedge Option 1

A natural hedge produces and sells locally with operating profits per state:

1. $1000M in U.S. + €500M in Europe at $1/€ = $1,500M
2. $500M in U.S. + €1,000M in Europe at $2/€ = $2,500M

The expected profit is $2,000M, with a variability risk of ± $500M.

Hedge Option 2

A natural hedge combined with a financial hedge that sells 500M future euros for $1.50 per euro (the expected financial return is zero and we neglect small transaction costs). The operating profits per state are:

1. $1,000M (U.S.) + €500M (Europe) + $750M − €500M (future) = $1,750M
2. $500M (US) + €1,000M (Europe) + $750M − €500M (future) = $1,250M + €500M at $2/€ = $2,250M

The expected profit is again $2,000M but with a reduced risk of ± $250M.

Hedge Option 3

An active operational hedge using allocation flexibility: We only produce in the low cost location—in Europe in state 1 and in the U.S. otherwise. The operating profits in each state are:

1. Sales: \$2,000M in U.S. + €1,000M in Europe. Cost: € 1,500M in Europe. Net = \$2,000M − €500 at \$1/€ = \$1,500M
2. Sales: \$1,000M in U.S. + €2,000M in Europe. Cost: \$1,500M in U.S. Net = \$500M + €2,000M at \$2/€ = \$3,500M

The expected profit is \$2,500M, an increase in value of 25% over the passive hedges! Recall that this option should at least require quick response in production (decide where to produce after the exchange rate is observed), which will likely be more costly than Hedge Options 1 and 2.

Hedge Option 4

The active operational hedge of Option 3 combined with a financial hedge would yield an expected profit of \$2,500M with reduced variance.

2.8.4 DIFFERENCES BETWEEN FINANCIAL AND OPERATIONAL HEDGING

A firm can simultaneously use both financial and operational hedging. For example, Ding et al. (2007) study integrated operational and financial hedging decisions faced by a global firm who sells to both home and foreign markets. Production occurs either at a single facility located in one of the markets or at two facilities, one in each market. The firm can use financial currency forward contracts to hedge currency risk. To further mitigate currency and demand risk, it can use ex-post operational flexibility.

Sometimes, however, complementing operational hedging with financial hedging may not be possible. For example, the planning horizon for a production facility may exceed 10 years. While operational hedging can be used, it is unlikely that financial hedging is available over that time-horizon. Financial hedging of capacity is also problematic if there is no capacity futures market that can replicate the capacity's cash flows (a swap can always be constructed if a counter party is available).

Whether a company should use both financial and operational hedging is the topic of current academic research. The answer depends on the type of financial contract, the operational system, and the correlation between the underlying financial asset and the operational risk under consideration. With perfect correlation, operational flexibility and financial hedging can complement each other, as Example 2.5 illustrates. Yet the optimal amount of operational flexibility that a firm should invest in depends on whether it engages in financial hedging or not. Chod et al. (2009) show that financial hedging with linear contracts increases the desired level of operational flexibility, while option contracts decrease it.

2.9 Tailored Operational Hedging

Earlier we said that risk management is an integral part of operations strategy. In this section, we will illustrate how risk management interacts with resource decisions (capacity size and type) and sourcing decisions. Furthermore, we will demonstrate how some generic operational hedging strategies can be tailored to specific situations.

2.9.1 TAILORED NATURAL HEDGING AT AUTO CO.

To illustrate the concept of tailored hedging, let us analyze how to tailor pure diversification to a particular setting. Consider, for example, a company that faces correlated demand risk, and manufactures two products, each on its own dedicated line. The question is how to size the capacity portfolio to mitigate risk. Mean-variance analysis of profits provides an answer.

To illustrate mean-variance analysis of a capacity portfolio, consider the stylized Auto Co. example introduced in Chapter 5 Van Mieghem (2008), first in a risk-neutral setting. Auto Co. is introducing two car models, Afour and Bassat. The Afour commands the higher price and unit contribution margin of $2000 versus the Bassat's $1,000. Investing in capacity involves a significant fixed cost and a variable cost that increases with the installed capacity size. For simplicity, we will assume that the fixed cost is the same for either product and hence does not impact our technology strategy choice. However, the capacity cost per unit for an Afour dedicated line is $800, slightly greater than the $700 for the Bassat line. The key risk stems from demand uncertainty and Figure 2.14 shows the total demand forecast.

The profit mean and variance for an investment budget of $100 million can be calculated for the demand data using simulation-based optimization. (Easily implemented in a spreadsheet that is downloadable from www.vanmieghem.us). Figure 2.15 plots the results for $100 million investments that vary their allocation

Demand Forecast			
Scenario	Afour	Bassat	Probability
1	70	150	5%
2	70	200	10%
3	70	250	15%
4	100	150	10%
5	100	200	20%
6	100	250	10%
7	130	150	15%
8	130	200	10%
9	130	250	5%

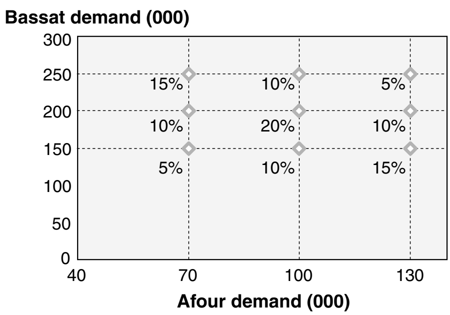


FIGURE 2.14 The demand forecast for Afours and Bassats. Uncertainty is captured via nine scenarios with associated probabilities. The forecast can be represented in a table (left) or a graph (right).

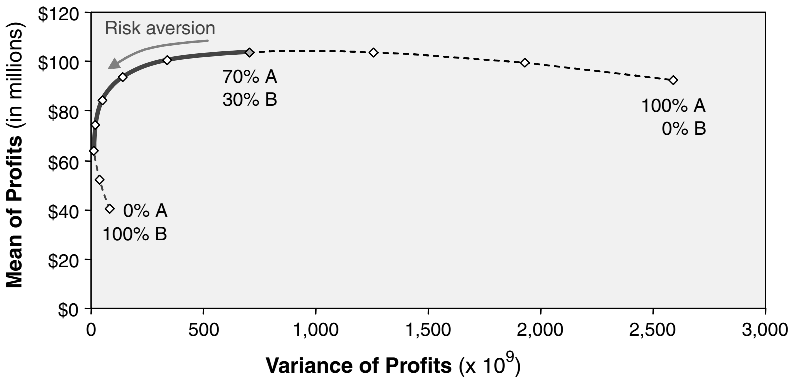


FIGURE 2.15 **Pure diversification results from serving two markets with dedicated resources A and B. The percentages show the relative resource investment for a given budget. A risk-averse manager can operationally hedge by rebalancing towards resource B, which serves the lower profit variance market.**

to Afour (A) and Bassat (B) capacity. A risk-neutral manager would maximize expected profits by investing $70 million in product line A and the remaining $30 million in B. In contrast, a risk-averse manager should move down the frontier (in bold in Fig. 2.15) and rebalance capacity towards B.

But why B? Given that market A's demand has a standard deviation of 30,000 with a unit contribution margin of $2,000, the standard deviation of its (budget-unconstrained) contribution is $60 million. Compare this with the $50 million for market B, whose demand has a standard deviation of 50,000 with a unit contribution margin of $1,000.

The general insight gained here is that firms can tailor their operational hedge by rebalancing dedicated capacities towards the resource that serves the market with lower profit variance. The Auto Co. example shows that this doesn't need to be the market with the lowest demand risk or the highest contribution margin. Rather, it is the product of these two factors that counts. The effectiveness of natural hedging increases as the pooled risks become more similar in magnitude and more negatively correlated. Indeed, a perfect zero-variance hedge would be obtained if both markets had equal profit variances and were perfectly negatively correlated.

2.9.2 TAILORED REDUNDANCY AND DYNAMIC POOLING WITH ALLOCATION FLEXIBILITY AT AUTO CO.

To illustrate active operational hedging, continue considering the Auto Co. example, enriched with two additional options. First, the firm can borrow investment funds, meaning that it has no budget constraint. Second, the firm can not only invest in the two dedicated resources, but also in a product-flexible line. The capacity portfolio now consists of three assets. The flexible line has higher capacity

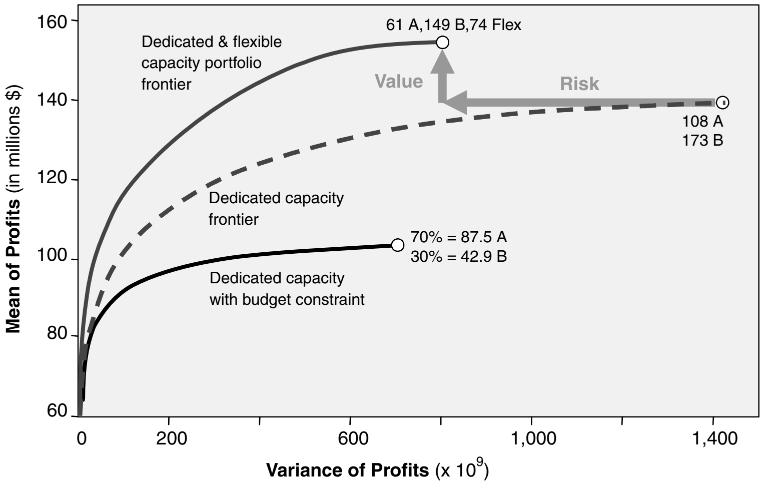


FIGURE 2.16 Adding flexibility not only mitigates profit variance risk through pooling (frontier is to the left of the dedicated capacity frontier) but also increases value by exploiting upside variations through contingent revenue maximization (frontier is above dedicated capacity frontier).

investment costs—the flexible line has the same fixed investment cost as a dedicated line but costs $900 per unit of annual capacity—but pools *and* exploits demand uncertainty. Given that flexible capacity serves as a substitute to the dedicated resources, it can also be interpreted as a form of adding reserves in the form of adding *redundancy*.

Figure 2.16 shows the magnitude of risk mitigation and the value enhancement of hedging with operational flexibility compared to pure diversification with dedicated assets. The system with the $100 million investment budget is dominated by relaxed budget constraints: Mean profits and profit variance risk increase, thus reflecting higher investments (108,000 A and 173,000 B annual car capacity versus 87,500 A and 42,900 B). In contrast, adding the option of investing in an additional flexible line here cuts profit variance risk roughly by 50% while increasing value by more than 10%. This shows that flexibility is attractive even to risk-neutral investors.

Risk-averse investors can further tailor the optimal operational hedge by rebalancing the capacity portfolio in two directions, as suggested by Figure 2.17 and studied by Van Mieghem (2007): to do so, they must increase the shares of the flexible capacity and of the resources serving the lower profit variance market (B). The latter reflects the pure diversification effect inherent in pooling, while the former demonstrates the profit variance risk mitigation power of flexibility (notice that the operational hedge can be so powerful that a risk-averse manager may even increase capacity relative to the risk-neutral levels). The tailored capacity balance depends on the manager's coefficient of risk aversion, as well as on the demand and processing network data.

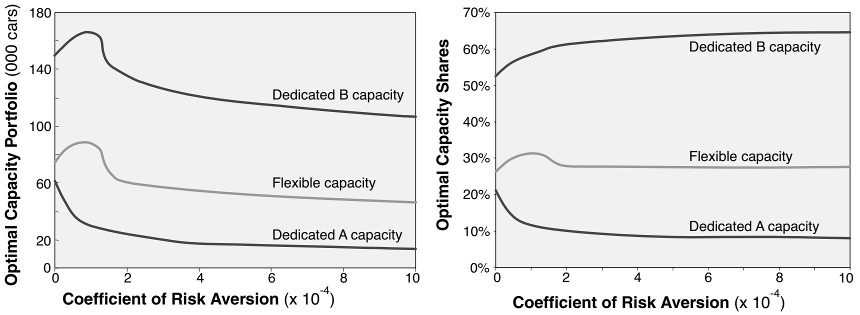


FIGURE 2.17 **Risk-averse managers tailor the operational hedge by rebalancing the capacity portfolio toward (1) more flexibility and (2) resource B, which serves the lower profit variance market.**

2.9.3 TAILORED OPERATIONAL HEDGING: BASE DEMAND, TAIL-POOLING, AND CHAINING

The appropriate capacity mix between flexible and dedicated capacity illustrates another tailoring dimension. Tailored flexibility serves mostly the uncertain part of the demand distribution (also known as "tail risk"), while most of the predictable "base demand" is allocated to dedicated resources. Benetton provides another example: Garment production of its base demand is allocated to a set of efficient subcontractors up to two quarters ahead of the season. Flexible in-house capacity produces garments quickly, thereby minimizing demand risk.

Tailored flexibility also works in service operations. Service representatives may be mostly dedicated to a certain product or region (base demand). As long as the resource-product allocations form a chain, service representatives can help out colleagues who are overloaded. Pooling benefits accrue while specialization benefits are enjoyed the majority of time. Bassamboo et al. (2009) show that this "tailored chaining" can outperform the chaining of only bi-flexible resources (first studied by Jordan and Graves [1995]) by balancing specialization (favoring dedicated resources) and pooling (favoring flexible resources) benefits.

Temporal tailoring of scale flexibility allocates quick response capacity to peak demand. Electricity capacity illustrates temporal tailoring in a single product setting: nuclear power serves base demand continuously while various levels of fossil fueled generators (including even jet generators) pick up peak demand.

2.9.4 TAILORED REDUNDANCY AND MULTISOURCING FOR SUPPLY AND PROJECT RISK

Multisourcing is a powerful strategy to mitigate supply failure risk. Consider the U.S. flu vaccine supply problem in 2004, when a major supplier (Chiron) was forced to close down due to violations of regulatory quality standards. The U.S. had roughly split the majority of its expected need of 100 million flu vaccines over

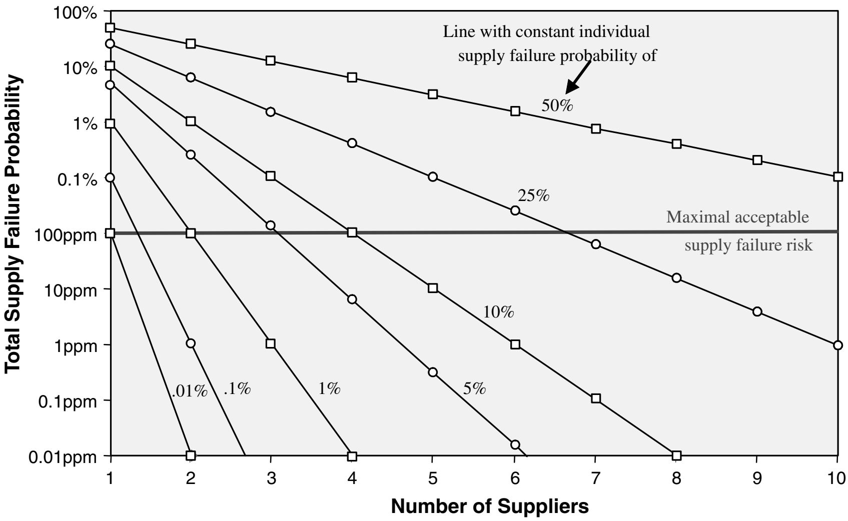


FIGURE 2.18 Tailored redundancy uses the appropriate number of suppliers depending on the maximal acceptable supply failure risk (100 parts per million here) and the failure probability of an individual supplier.

Chiron and one other supplier. Because of the long lead times (about 8 months), it had little recourse; thus, flu vaccines were put on allocation, causing a serious political outcry. This was in marked contrast with the U.K.'s hedged strategy, which uses six suppliers for a target demand of only 14 million.

Tailored redundancy selects the appropriate number of suppliers based on the maximal acceptable supply failure risk and the failure risk profiles of individual suppliers. For example, consider the simplest situation where supply risk is all or nothing (similar perhaps to Chiron's flu vaccine problem) and the failure probability p is the same for all suppliers and independent of one another. Supply totally fails only when all suppliers fail. The probability of total supply failure when using N suppliers thus is p^N. Figure 2.18 shows this relationship in a log-linear plot. This determines the minimum number of suppliers needed to diversify supply risk below a maximal tolerable level. Clearly, more suppliers are needed if they are more unreliable or if maximal acceptable risk levels are tighter.

This insight extends to the setting where supply failure is manifested by an uncertain or *random yield* (or probability that a unit ordered is of acceptable quality) and can differ from supplier to supplier. The analysis is much more involved and has only recently been done by Federgruen and Yang (2007). They found that total supply should be allocated to a tailored number of suppliers, each supplier's allocation being proportional to the mean-to-variance ratio of that supplier's yield distribution. That allocation scheme also minimizes variable sourcing costs.

Redundancy through common platforms or even parallelism can also mitigate project risk. For example, when Toyota develops a new car, it often produces a large number of prototypes, several of them in parallel (Sobek II et al. 1999). It decides which type will eventually be commercialized as close to market introduction as possible, in order to have the product better respond to market needs. While redundancy increases the costs of the R&D stage, it gives Toyota an option to significantly increase project revenues by commercializing the most profitable prototype.

2.10 Guidelines for Operational Risk Management

2.10.1 IMPLEMENT AN OPERATIONAL RISK MANAGEMENT PROCESS

In most companies, risk management is the responsibility of the CFO. In addition to financial risks, companies should also acknowledge, identify, and assess operational risks. Setting up a formal operational risk management process under a senior operations manager is a necessary first step. For example, Grainger has a "business continuity department" of about 15 people that anticipates, evaluates,and mitigates operational risks.

2.10.2 USE A MULTIFACETED APPROACH TAILORED TO THE TYPE OF RISK AND PRODUCT LIFE CYCLE

No single size fits all. Risk mitigation should use the right mix of multiple financial and operational hedging strategies, depending on the type of risk. For example, supply risk of short life cycle products is best mitigated with supplier diversification and demand management techniques such as contingent substitution and pricing. For long life cycle products, inventory, contingent supply, and continuous risk monitoring of suppliers may be more appropriate. Make the distinction between intermittent and recurrent supply risks.

Not only the length, but also the stage in the product life cycle determines appropriate tactics. Technical innovation risk in the pharmaceutical industry is mitigated by redundancy (developing several designs in parallel), faster and earlier drug trials (testing), and retaining flexibility so important decisions can be postponed (e.g., by using modular facility construction).

2.10.3 USE A PORTFOLIO APPROACH

While each risk needs a tailored response, remember that the organization's total risk exposure enjoys portfolio diversification benefits. Such a portfolio approach often justifies investment in redundant assets or more expensive flexible assets.

2.10.4 REALIZE THAT OPERATIONAL HEDGING MAY INCUR ADDITIONAL COSTS

There is no such thing as a free lunch. The benefits of operational hedging may involve additional hidden costs. For example, multilocation processing incurs a loss of scale, requires procurement from a wider supply base, slows down the learning curve process, and may produce less-consistent quality. Good risk management tries to reduce these costs over time.

2.10.5 REDUCING RISK IS MORE POWERFUL THAN MITIGATING EXPOSURE

In the long run, reducing and eliminating sources of risk is often more profitable than mitigating their impact with fences, counterbalancing actions, or band-aids. For example, exposure to demand uncertainty can be mitigated through pooling and reserves like safety inventory or capacity. Yet, initiatives like lead time reduction, postponement, quick response, better forecasting, and information sharing reduce the demand uncertainty (and with it, the need for mitigation).

Operations management has a rich heritage in eliminating the root cause of "problems" as illustrated by the success of the Toyota Production System and continuous improvement programs such as lean operations, Six Sigma, and total quality management. Such operations improvement programs should be an important component of any risk management program.

REFERENCES

Bassamboo A., Randhawa R.S., Van Mieghem J.A. 2009. A little flexibility is all you need: Asymptotic optimality of tailored chaining and pairing in queuing systems. Working Paper, Northwestern University.

Chod J., Rudi N., Van Mieghem J.A. 2010. Operational flexibility and financial hedging: Complementarity or substitutes? *Management Science*, **56**(6):1030–1045.

Chopra S., Sodhi M.S. 2004. Managing risk to avoid supply chain breakdown. *Sloan Management Review*, Fall 2004.

Ding Q., Dong L., Kouvelis P. 2007. On the integration of production and financial hedging decisions in global markets. *Operations Research*, **55**(3):470–489.

Dong L., Liu H. 2007. Equilibrium forward contracts on nonstorable commodities in the presence of market power. *Operations Research*, **55**(1):128–145.

Federgruen A., Yang N. 2008. Selecting a portfolio of suppliers under demand and supply risks. *Operations Research*, **56**(4):916–936.

Gaur V., Seshadri S. 2005. Hedging inventory risk through market instruments. *Manufacturing and Service Operations Management*, **7**(2):103–120.

Jordan W., Graves S. 1995. Principles on the benefits of manufacturing process flexibility. *Management Science*, **41**:557–594.

Latour A. 2001. Trial by fire: A blaze in Albuquerque sets off major crisis for cell-phone giants—Nokia handles supply shock with aplomb as Ericsson of Sweden gets burned—Was Sisu the difference? *Wall Street Journal*, January 29.

National Interagency Fire Center. 2002. Interagency helicopter operations guide. www.nifc.gov/ihog/pdf/chapters/ch03.pdf.

Pindyck R.S., Rubinfeld D.L. 1989. *Microeconomics*. New York: Macmillan.

Sharpe W. 1985. *Investments*. 3rd ed. Prentice-Hall: Englewood Cliffs, NJ.

Sheffi Y. 2005. *The Resilient Enterprise*. The MIT Press.

Sobek II D.K., Ward A.C., Liker J.K. 1999. Toyota's principles of set-based concurrent engineering. *Sloan Management Review*, **40**:67–83.

Tomlin B. 2006. On the value of mitigation and contingency strategies for managing supply-chain disruptions risks. *Management Science*, **52**(5):639–657.

Van Campenhout M. 2004. Private correspondence on "Mechelen's position on handling the consequences of the Jackson tornado". October 22.

Van Mieghem J.A. 2007. Risk mitigation in newsvendor networks: Resource diversification, flexibility, sharing, and hedging. *Management Science*, **53**(8):1269–1288.

Van Mieghem J.A. 2008. *Operations Strategy: Principles and Practice*. Charlestown, Massachusetts: Dynamic Ideas. Available from www.dynamic-ideas.com.

CHAPTER THREE

The Effect of Supply Chain Disruptions on Corporate Performance

KEVIN B. HENDRICKS AND VINOD R. SINGHAL

3.1 Introduction

Managers are becoming increasingly aware that their company's reputation, earnings consistency, and ability to deliver better shareholder returns are increasingly dependent on how well they manage supply chain disruptions. Although firms have always faced the risk of supply chain disruptions, the attention it receives has increased dramatically in recent years. This is likely driven by at least four developments. First, supply chains have become more complex due to globalization, outsourcing, single sourcing, and the focus on removing slack from supply chains. While many of these strategies have improved performance, these strategies have also made supply chains more prone to disruptions.

Second, the focus on supply chain disruptions has increased following a number of costly and highly publicized supply chain disruptions. National and

Handbook of Integrated Risk Management in Global Supply Chains, First Edition. Edited by Panos Kouvelis, Lingxiu Dong, Onur Boyabatli, and Rong Li.

local media are filled with news reports on the increase in supply chain disruptions, and the fact that many companies are unable to cope with these disruptions. Some recent examples, include the disruptions due to Mattel's recall of 21 million toys due to safety issues (Casey and Pasztor 2007); Boeing's unexpected delay in introducing its much anticipated 787 Dreamliner because of difficulties in coordinating global suppliers (Lunsford 2007); and recall of contaminated meat, pet foods, and pharmaceuticals products (Kesmodel 2008 and Fairclough 2008).

Third, academicians and practitioners are discussing the impact of supply chain disruptions on performance as well as highlighting the need to adopt practices that can prevent disruptions (Kilgore 2003, Radjou 2002, Cachon and Lariviere 2001, Lakenan et al. 2001, Lee et al. 1997, and Fisher 1997). A survey by FM Global of more than 600 financial executives finds that supply chain risks pose the most significant threat to profitability (Smyrlis 2006). A survey by Accenture of 151 supply chain executives finds that 73% indicate that their firms experienced supply chain disruptions in the past five years (Ferrer et al. 2007). Various studies identify drivers of supply chain risk, and develop frameworks and strategies for managing and mitigating supply chain risk (Fisher 1997, Lee et al. 1997, Chopra and Sodhi 2004, Billington et al. 2002, Sheffi 2005, Kleindorfer and Said 2005, Tang 2006, Tomlin 2006, and Craighead et al. 2007).

Finally, the passage of the Sarbanes-Oxley Act of 2002 makes senior executives more responsible for forecasts of performance and protection of shareholder value. This has heightened the need to identify and manage various risks, including supply chain disruptions.

This chapter addresses three issues that are critical in managing supply chain disruptions. First, it summarizes evidence from recent empirical research on the financial consequences of supply chain disruptions (Hendricks and Singhal 2003, 2005a, 2005b). One of the reasons why many companies are not adequately prepared for responding to supply chain disruptions is that they do not have a good understanding of the magnitude and persistence of the negative consequences of disruptions on financial performance. While anecdotes make for splashy headlines, they do not provide the objective evidence that many senior executives are looking for to better understand the financial consequences of supply chain disruptions to make decisions about the initiatives and investments they should undertake to manage disruptions. The financial consequences are examined by documenting the impact of supply chain disruptions on shareholder returns, share price volatility, and profitability. Second, it offers insights into the factors that can increase the chances of disruptions to guide managers as they assess the chances of disruptions. Third, it highlights some of the strategies and practices in managing disruptions using examples from Wal-Mart, Mattel, and Boeing.

The evidence and discussion presented in this chapter is important for a number of reasons. As mentioned above, it fills a gap in the literature regarding the financial consequences of demand-supply mismatch. Supply chain disruptions are a form of demand-supply mismatches. Although the conventional belief is that supply-demand mismatch will have negative financial consequences, there is very little rigorous empirical evidence on the magnitude and severity of the financial consequences.

Efficiency, reliability, and responsiveness of supply chains are key drivers of a firm's profitability. Kilgore (2003) and Radjou (2002) suggest that much of the supply chain management efforts in the recent past have focused on increasing the efficiency (lowering costs) of supply chain operations, and less on increasing the robustness and reliability of supply chains. This could partly be because unlike efficiency, it is much harder to place a value on robustness and reliability. Disruptions are an indication that a firm's supply chain is not reliable and robust. By associating disruptions with financial outcomes, we provide an estimate on the value of reliable and robust supply chain performance.

This chapter also adds to the recent research that has begun to quantify the impact of supply chain management strategies and practices on operating performance. One stream of research has focused on developing mathematical models of supply chain issues to understand how alternate ways of managing supply chains affect capital costs, operating costs, inventories, and service levels (see, for example, Barnes-Schuster et al. 2002, Milner and Kouvelis 2002, Taylor 2002, Aviv 2001, Cachon and Lariviere 2001, and Cachon and Fischer 2000). Another stream of research has attempted to empirically establish the relationship between supply chain practices and performance. The approach used is to develop conceptual and theoretical frameworks of the drivers of supply chain performance, identify supply chain practices, use surveys to measure the intensity with which these practices are implemented, and link these to performance changes reported by survey respondents (Rosenzweig et al. 2003, Frohlich and Westbrook 2001, Narasimhan and Jayaram 1998, and Shin et al. 2000). Although significant research has been done on the relationship between supply chain performance and financial performance, most of the existing evidence is based on hypothetical or self-reported data. Hence, it is not clear how well the evidence correlates to actual performance.

The next section describes the sample, performance metrics, and methodology for estimating the financial impacts. Section 3.3 presents results on the impact of supply chain disruptions on shareholder value, share price volatility, and profitability. Section 3.4 discusses the various drivers of supply chain disruptions. Section 3.5 discusses what firms can do to reduce the frequency of disruptions and mitigate the negative consequences of disruptions. This section also discusses the examples of Wal-Mart, Mattel, and Boeing in dealing with disruptions and what can be learned from their experiences. The final section summarizes the chapter.

3.2 Sample, Performance Metrics, and Methodology

The evidence presented in this report is based on an analysis of more than 800 supply chain disruptions that were publicly announced from 1989 to 2001. These announcements appeared in *The Wall Street Journal* and/or the Dow Jones News Service, and were about publicly traded companies that experienced production or shipping delays. Some examples of such announcements are:

- *"Sony Sees Shortage of Playstation 2s for Holiday Season",* The Wall Street Journal, *September 28, 2000.* The article indicated that because of component shortages, Sony has cut in half the number of PlayStation 2 machines it can manufacture for delivery.
- *"Motorola 4th Quarter Wireless Sales Growth Lower Than Order Growth",* The Dow Jones News Service, *November 18, 1999.* In this case Motorola announced that its inability to meet demand was due to the shortage of certain types of components and that the supply of these components is not expected to match demand sometime till 2000.
- *"Boeing Pushing for Record Production, Finds Parts Shortages, Delivery Delay,"* The Wall Street Journal, *June 26, 1997.* The article discusses reasons for the parts shortages, the severity of the problems, and the possible implications.
- *"Apple Computer Inc. Cuts 4th period Forecast Citing Parts Shortages, Product Delays",* The Wall Street Journal, September 15, 1995. Apple announced that earnings would drop because of chronic and persistent part shortages of key components and delays in increasing production of new products.

The performance effects of the above mentioned instances of supply chain disruptions are estimated by examining performance over a three-year time period starting one year before the disruption announcement date and ending two years after the disruption announcement date. Two stock market based metrics are used in the analysis:

- Shareholder returns as measured by stock returns that include changes in stock prices as well as any dividends declared.
- Share price volatility.

The effect of disruptions on profitability is examined using the following measures:

- Operating income (sales minus cost of goods sold minus selling and general administration).
- Return on sales (operating income divided by sales).
- Return on assets (operating income divided by total assets).
- Costs (sum of cost of goods sold and selling and general administration cost).
- Total assets.
- Total inventory.

To control for industry and economy affects that can influence changes in the above performance measures, the performance of the disruption-experiencing firms is compared against benchmarks of firms that are in the same industry with similar size and performance characteristics. Appendix 3.1 briefly describes the methodology used for estimating stock returns, share price volatility, and profitability changes attributed to disruptions.

3.3 The Effect of Supply Chain Disruptions on Corporate Performance

3.3.1 THE EFFECT OF SUPPLY CHAIN DISRUPTIONS ON SHAREHOLDER VALUE

Figure 3.1 depicts the shareholder value effects on the day supply chain disruption is publicly announced. The effects that can be attributed to a disruption is estimated by comparing the stock returns of disruption-experiencing firms against four different benchmarks that serve to control for normal market and industry influences on stock returns. All shareholder value effects in Figure 3.1 are significantly different from zero at the 1% level.

The evidence indicates that supply chain disruptions are viewed very negatively by the market. On average shareholders of disruption-experiencing firms lose:

- 7.18% relative to the benchmark that consists of the portfolio of all firms that have similar prior performance, size, and market to book ratio of equity to the disruption-experiencing firm (portfolio matched benchmark).
- 7.17% relative to the firm that has similar prior performance and market to book ratio of equity, and is closest in size to the disruption-experiencing firm (size matched benchmark).

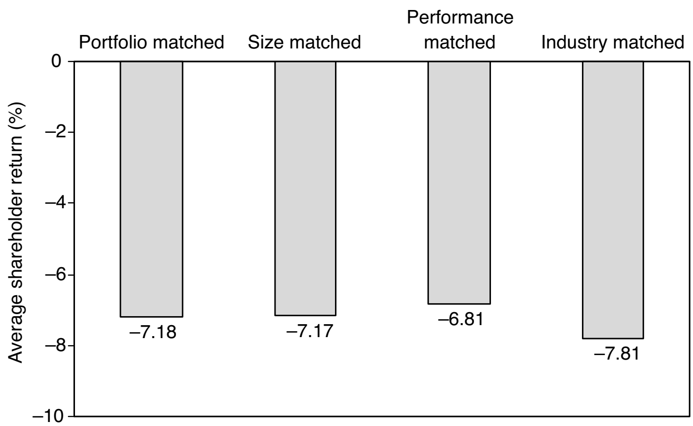


FIGURE 3.1 The average shareholder return on the day information about a disruption is publicly announced. Portfolio, size, performance, and industry matched are different set of benchmarks used to estimate the relative stock price performance of the firms that experience disruptions.

- 6.81% relative to the firm that has similar size and market to book ratio of equity, and is closest in terms of prior performance to the disruption-experiencing firm (performance matched benchmark).
- 7.81% relative to the firm that has similar size, prior performance, and market to book ratio of equity, and is closest in terms of the industry to the disruption-experiencing firm (industry matched benchmark).

When one examines the relative stock price performance during the periods before and after the disruption announcement, the shareholder value effects are much worse than those depicted in Figure 3.1. Figure 3.2 depicts the stock price performance starting one year before and ending two years after the disruption announcement date. The stock price performance is measured relative to the portfolio of all firms that have similar prior performance, size, and market to book ratio of equity to the disruption-experiencing firm (i.e., portfolio matched).

During the year before the disruption announcement, stocks of disruption-experiencing firms underperformed their benchmark portfolio by nearly 14%, significantly different from zero at the 1% level. Given that announcements are acknowledgement of disruptions that have already occurred and firms have an incentive to delay the acknowledgement, one can see why disruptions can be partially anticipated. Indication of disruptions could show up in many ways including difficulty in obtaining the firm's products, build-up of inventories at suppliers, the firm's inventory falling below critical levels, press releases by other supply chain partners, articles in business press, and analyst research reports. Accordingly, the market may have assigned a probability that the firm is likely to suffer from a disruption, and hence may have incorporated part of the economic impact of disruptions in stock prices even before the disruption announcement.

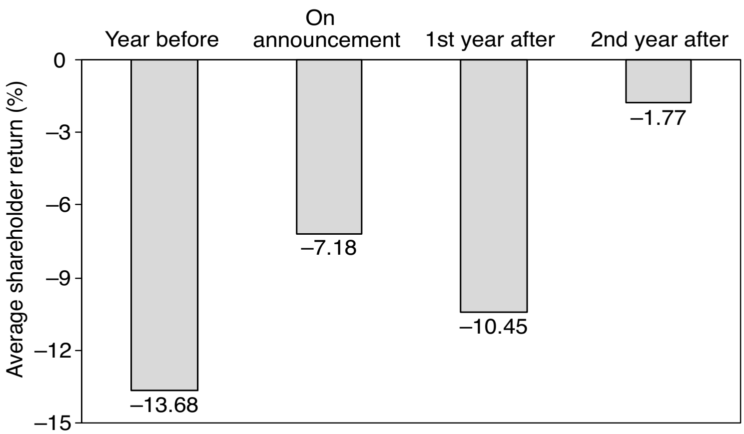


FIGURE 3.2 The average shareholder returns during the year before the disruption announcement, on announcement, and each of the two years after the disruption announcement. The shareholder returns are estimated relative to the portfolio of all firms that have similar prior performance, size, and market to book ratio of equity to the disruption-experiencing firm.

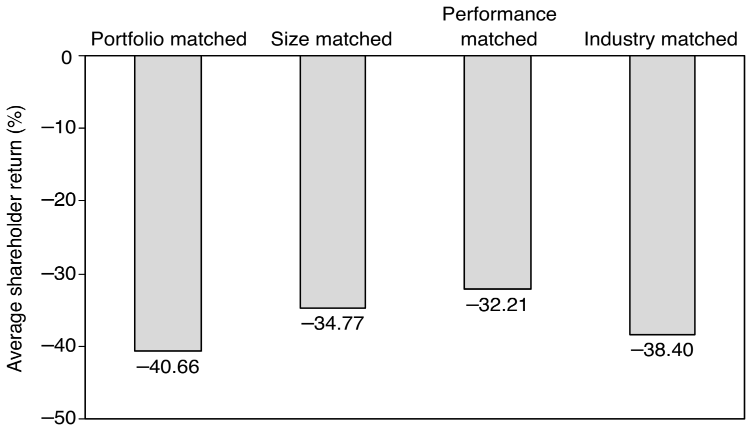


FIGURE 3.3 **The average shareholder returns relative to various benchmarks measured over a three year period that begins a year before the disruption announcement and ends two years after the disruption announcement. Portfolio, size, performance, and industry matched are different set of benchmarks used to estimate the relative stock price performance of the firms that experience disruptions.**

Even after the announcement of disruptions, firms continue to experience worsening stock price performance. In the year after the disruption announcement firms on average lose another 10.45% relative to their benchmark portfolios, significantly different from zero at the 1% level. Although the negative trend continues in the second year after disruption, the magnitude of underperformance of 1.77% (insignificantly different from zero) is not as high as that during the year before and the first year after the disruption announcement. More importantly, the results show that firms do not recover during this period from the negative stock price performance that they experienced in the prior two years, indicating that the loss associated with disruptions is not a short-term effect.

Figure 3.3 depicts the extent of shareholder value loss associated with disruptions over the three-year period. Depending on the benchmark used, the average level of underperformance on shareholder returns ranges from 33% to 40%, significantly different from zero at the 1% level. One way to judge the economic significance of this level of underperformance is the fact that on average stocks have gained 12% annually in the last two decades. Even if a firm experiences one major supply chain disruption every 10 years, the annual return would be close to 8% or 9%, which is a significant difference when one takes into account the effect compounding over long periods. Clearly, it pays to avoid supply chain disruptions. These results also underscore the importance of why senior executives must be aware of and actively involved in monitoring and managing the performance of their firm's supply chain.

The average level of share price underperformance documented in Figure 3.3 is not driven by a few outliers or special cases. Figure 3.4 shows that anywhere from 62% to 68% of the firms that experience disruption underperform their

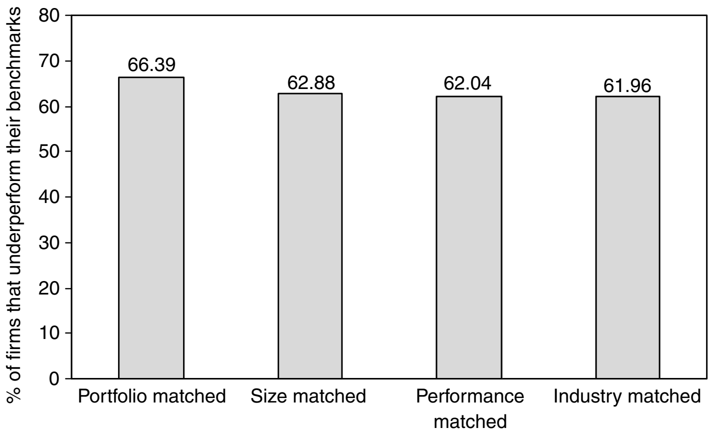


FIGURE 3.4 The percent of disruption experiencing firms that underperform their benchmarks over a three-year period that begins a year before the disruption announcement and ends two years after the disruption announcement. Portfolio, size, performance, and industry matched are a different set of benchmarks used to estimate the relative stock price performance of the firms that experience disruptions.

respective benchmarks over a three-year period, which is a statistically significant level of underperformance.

In summary, Figures 3.1 through 3.4 indicate the following:

- Supply chain disruptions result in significant short-term and long-term shareholder value losses; 33% to 40% stock price underperformance over three years is both economically and statistically significant.
- Firms that experience disruptions do not recover quickly from the stock price underperformance. Disruptions have a long-term devastating effect on shareholder value.

3.3.2 THE EFFECT OF SUPPLY CHAIN DISRUPTIONS ON SHARE PRICE VOLATILITY

Supply chain disruptions can create uncertainty about a firm's future prospects and can raise concerns about its management capability as disruptions indicate management inability to manage and control crucial business processes. Disruptions may also lead to questions and concerns about a firm's business strategy. Disruptions could therefore increase the overall risk of the firm. Understanding how disruptions can affect the risk of the firm is important for a number of reasons:

- Risk is a critical factor used by investors to value a firm's securities. Risk influences the return that investors demand for holding securities and hence directly affects the pricing of securities.

- The discount rate used in capital budgeting is directly related to the risk of the firm. Furthermore, the cost of capital when raising capital via equity and/or debt is influenced by the risk of the firm. The higher the risk, the higher is the cost of capital.
- Increased risk can make the firm's shares a less attractive currency for acquisitions as potential targets may be less willing to do deals that depend on volatile share prices.
- Rating agencies such as Moody's and S&P 500 consider the risk of the firm in determining a firm's credit rating. Increase in risks can result in downgrading of debt by credit rating agencies, making it more expensive and difficult to raise capital. It can also increase the probability of financial distress as the chances of the firm not being able to cover its fixed commitments increase as the risk increases.
- Risk changes can create conflicts between the various stakeholders. An increase in share price volatility transfers wealth from bondholders to shareholders, a potential source of conflict that may require management time and attention. Risk-averse employees may demand higher compensation to work for a firm that has high risk. Suppliers and customers may also be wary of dealing with the firm that has high risk and may demand some form of assurances and guarantees before doing business with the firm, thereby raising the cost of doing business for the firm.

To estimate the effect of disruptions on risk, this study compared the share price volatility before and after the disruption announcement date. Share price volatility is measured by the standard deviations of stock returns, which are estimated annually for four years, starting two years before through two years after the disruption announcement. To control for other factors that could affect volatility, percent changes in the standard deviation of stock returns of the disruption-experiencing firms are compared against that of a matched control sample.

Figure 3.5 gives share price volatility (standard deviation of stock returns) using daily stock returns for the firms that experienced supply chain disruptions. In computing the volatility, we exclude returns over a trading period of 21 days that starts 10 trading days before the announcement day and ends 10 trading days after the announcement day. The figure indicates that the share price volatility is monotonically increasing starting two years before the disruption announcement and ending two years after the disruption. For example, the standard deviation of stock returns in the second year before the disruption announcement was 4.13% and since then has steadily increased to 5.05% in the second year after the disruption announcement. The evidence supports the view that disruptions increase the share price volatility, and hence the risk of the firm.

One can get a better idea of the extent of share price volatility changes by comparing the change in the share price volatility of disruption-experiencing firms against the change in share price volatility experienced by a control sample. Figure 3.6 reports these results. The results indicate that after adjusting for other factors that could affect share price volatility there is still a significant increase

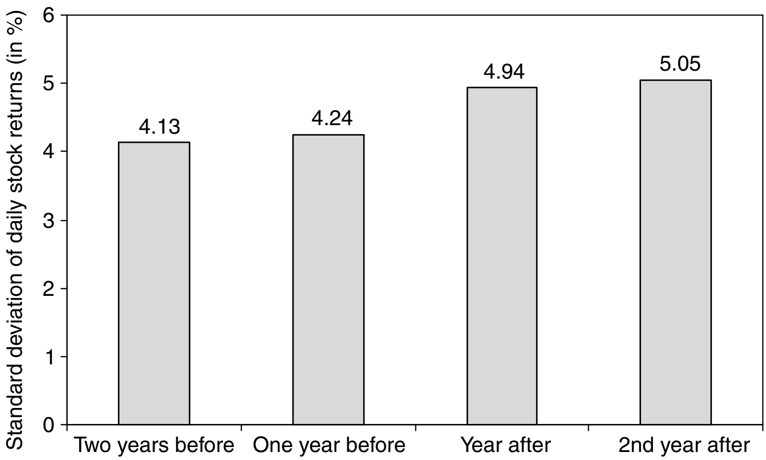


FIGURE 3.5 **Estimated standard deviation of stock returns over a four-year time period for the sample of firms that experienced disruptions.**

in volatility that can be attributed to the disruption. Much of this increase happens after the disruption announcement. For example, the share price volatility increases by 13.5% (significantly different from zero at the 1% level) in the year after the disruption when compared to the volatility one year before the disruption announcement. Furthermore, the share price volatility remains at this high level for at least the next year or two. Overall, disruptions increase the risk of the firm.

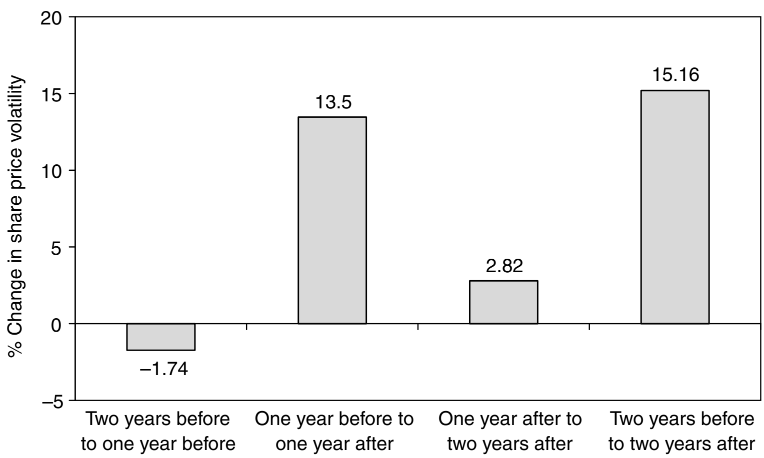


FIGURE 3.6 **Estimated percent changes in standard deviation of stock returns over a four-year time period. The reported percent changes are the difference between the percent changes of the disruption-experiencing firms and its control firms.**

3.3.3 THE EFFECT OF SUPPLY CHAIN DISRUPTIONS ON PROFITABILITY

The magnitude of stock price underperformance associated with supply chain disruptions and the lack of any recovery may surprise many and could raise the issue whether the significant stock price underperformance is due to a corresponding reduction in profitability or it is simply a matter of stock market overreaction. This issue is explored by documenting the long-term effects of disruptions on operating income, sales growth, cost growth, as well as changes in the level of assets and inventories. As in the case of the analysis of stock price performance, profitability effects are estimated starting one year before and ending two years after the disruption announcement.

The key results of this analysis are highlighted in Figures 3.7 through 3.9. To control for industry, economy, and others affects, the performance of the disruption-experienced firms is compared to controls using the three different control samples described in Appendix 3.1. Since the three control samples give similar results, the results from the control sample where most of the sample firms are matched are reported. Since accounting data are more prone to extreme values or outliers, the average values reported are those obtained after trimming 1% on each tail. The median changes, which are less influenced by outliers, are also reported.

The results indicate that supply chain disruptions have a devastating effect on profitability. Figure 3.7 shows that firms that experience disruptions on average experience a 107% decrease operating income, 114% decrease in return on sales, and 92% decrease in return on assets, all significantly different from zero at the 1% level. Outliers are not driving the negative mean changes in operating-income based measures. The median of the percent changes in operating income, return on sales, and return on assets are −42%, −32%, and −35%, respectively, all significantly different from zero at the 1% level.

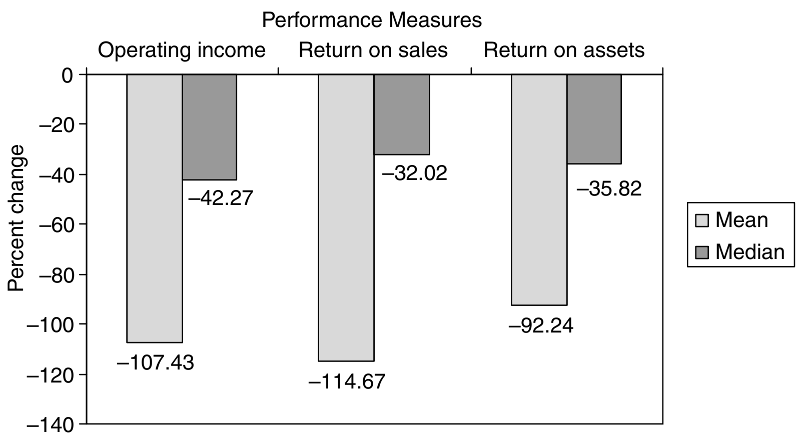


FIGURE 3.7 Control-adjusted changes in profitability related measures from supply chain disruptions. Performance effects are estimated starting one year before and ending two years after the disruption announcement.

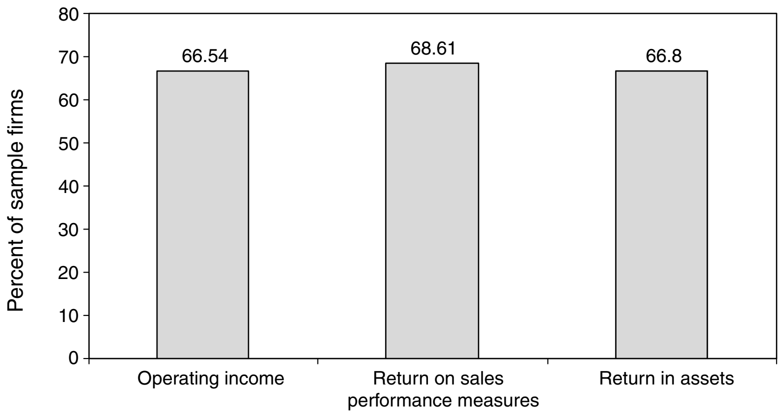


FIGURE 3.8 **The percent of disruption experiencing firms that underperform their benchmarks. Performance effects are estimated starting one year before and ending two years after the disruption announcement.**

The proportion of firms experiencing negative performance (see Figure 3.8) indicates that disruptions are bad news across the board. For example, nearly 67% to 69% of the sample firms experienced a negative change in operating income.

Figure 3.9 indicates that supply chain disruptions negatively affect sales. The mean (median) percent change in sales is about −7% (−3%), significantly different from zero at the 1% level. Nearly 54% of the sample firms experienced

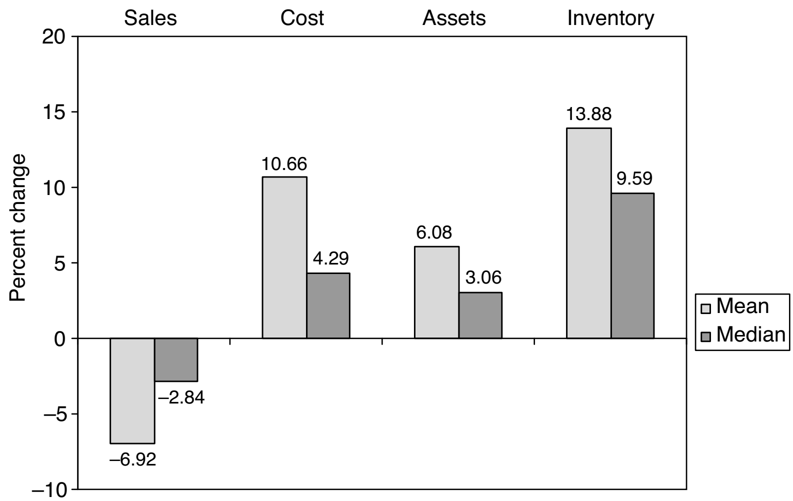


FIGURE 3.9 **Control-adjusted changes in sales, costs, assets, and inventories of firms experiencing supply chain disruptions. Performance effects are estimated starting one year before and ending two years after the disruption announcement.**

negative sales growth. Disruptions also increase total costs. The mean (median) change in total costs is about 11% (4%). Nearly 65% of the sample firms experience an increase in total costs, significantly different from zero at the 1% level. The drop in sales together with the increase in total costs explains the economically significant drop in operating income-based measures of Figure 3.7.

3.4 Drivers of Supply Chain Disruptions

The analysis of the effect of supply chain disruptions on financial performance is valuable because it provides firms with a perspective on the economic effect of poor supply chain performance. The evidence clearly indicates that ignoring the possibility of supply chain disruptions can have devastating economic consequences. As one reflects on this evidence, a natural question is "what are the primary drivers of supply chain disruptions?" Given the recent heightened awareness of the risk of supply chain disruptions many experts have offered insights into the factors that can increase the chances of disruptions. Some of these factors are discussed next with the intention that these factors can serve as a guideline for managers as they assess the chances of disruptions in their supply chains. The chances of experiencing disruptions are higher now and in the future than in the past because of some recent trends and practices in managing supply chains:

- **Competitive environment.** There is no doubt that most industries are facing a vastly different competitive environment today than a decade ago. Today's markets are characterized by intense competition, volatile demand, increased demand for customization, increased product variety, and short product life cycles. These trends are expected to intensify in the future. These conditions make it very challenging to match demand with supply. In particular, firms are facing increasing difficulty in forecasting demand and adjusting to unexpected changes in product life cycles and changing customer preferences.
- **Increased complexity.** The complexity of supply chains has increased due to global sourcing, managing large number of supply chain partners, the need to co-ordinate across many tiers of supply chains, and dealing with long lead times. This increased complexity makes it harder to match demand and supply, thereby increasing the risk of disruptions. The risk is further compounded when various supply chain partners focus on local optimization, when there is lack of collaboration among supply chain partners, and when there is lack of flexibility in the supply chain.
- **Outsourcing and partnership.** Increased reliance on outsourcing and partnering has heightened interdependencies among different nodes of the global supply networks and increased the chances that a disruption or problem in one link of the supply chain can quickly ripple through the rest of the chain, bringing the whole supply chain to a quick halt. While many experts have talked about the virtues of outsourcing and partnerships, for these to truly work well it is important that supply chain partners collaborate,

share information and plans, and have visibility in each other's operations. Such changes require major investments in connected information systems, changes in performance metrics, commitment to share gains, and building trust among supply chain partners, all of which are not easy to achieve.

- **Single sourcing.** Single sourcing strategies have reduced the purchase price and the administrative costs of managing the supplier base, but may have also increased the vulnerability of supply chains if the single-source supplier is unable to deliver on time.
- **Limited buffers.** Focus on reducing inventory and excess capacity and squeezing slack in supply chains has more tightly coupled the various links leaving little room for errors. Just-in-time delivery and zero inventory are commonly cited goals but without careful consideration of the fact that these strategies can make the supply chain brittle.
- **Focus on efficiency.** Supply chains have focused too much on improving efficiency (reducing costs). Firms are responding to the cost squeeze at the expense of increasing the risk of disruptions. Most firms do not seem to consider the inverse relationship between efficiency and risk. Strategies for improving efficiency can increase the risk of disruptions.
- **Over-concentration of operations:** In their drive to take advantage of economies of scale, volume discounts, and lower transaction cost, firms have over-concentrated their operations at a particular location, or with their suppliers or customers. Over-concentration reduces the flexibility of the supply chain to react to changes in the environment and leads to a fragile supply chain that is susceptible to disruptions.
- **Poor planning and execution.** Poor planning and execution capabilities result in more incidents of demand-supply mismatches. Plans are often too aggregate, lack details, and are based on inaccurate inventory and capacity information. Lack of good information systems hinders the ability of the organization to be aware of what is happening. Lack of forward-looking metrics affects the ability of firms to anticipate future problems and be pro-active in dealing with these problems. Firms also have limited visibility into what is happening in upstream and downstream supply chain partners. Most firms have limited abilities and capabilities to identify and manage supply chain exceptions. This is further compounded by the lack of synchronization and feedback between supply chain planning and supply chain execution.

3.5 What Can Firms Do To Mitigate the Chances of Disruptions?

There are no doubts that many of the above-mentioned practices and trends have led to improvements in supply chain performance and profitability. Nonetheless, they may have also contributed to supply chains becoming more susceptible and vulnerable to disruptions. The challenge therefore is to devise approaches that can deal more effectively with disruptions, while not sacrificing efficiency. Some of these approaches are briefly outlined below:

- **Improving the accuracy of demand forecasts.** One of the primary reasons for demand-supply mismatches is inaccurate forecasts. Bringing some quantitative rigor to forecasting can certainly help improve the accuracy and reliability of forecasts. Firms should consider not only the expected demand forecast but also the demand forecast error (variance) in developing plans. This would give planners an idea of what kind of deviation may happen from the mean value. Firms should also recognize that long-term forecasts are inherently less accurate than short-term forecasts as well as the fact that disaggregate forecasts are less accurate than aggregate forecasts. These considerations will enable planners to look more carefully at the forecasts they receive from sales and marketing. Forecasts often go bad when firms do not dynamically adjust forecasts, and fail to consider events outside their own organizations that could have a material effect on forecasts. Furthermore, firms often make forecasts assuming static lead times, transit time, capacity, and transportation and distribution routes. These assumptions must constantly be questioned to make adjustments as and when needed. Long planning time horizons that are frozen also makes it harder to develop accurate forecasts.
- **Integrate and synchronize planning and execution.** Firms have become sophisticated in their planning activities. But plans are often insulated from execution reality. In many cases plans are tossed over the wall for execution. Managers responsible for execution make adjustments to these plans to reflect current operating conditions. Such adjustments can grow over time but are seldom communicated to the planners, resulting in lack of integration between development and execution of plans. By better coordinating and integrating planning and execution many of the problems with supply-demand mismatches can be avoided.
- **Reduce the mean and variance of lead time.** Forecasting inaccuracy and disconnect between planning and execution can be particularly devastating when lead times are long and highly variable. Reducing the mean and variance of lead time can help reduce the level of uncertainties in the supply chain. Some of the following practices can help reduce the mean and variance of lead times:
  - Remove non-value added steps and activities
  - Improve the reliability and robustness of manufacturing, administrative, and logistics processes
  - Pay close attention to critical processes, resources, and material
  - Incorporate dynamic lead-time considerations in planning and quoting delivery times
- **Collaborate and cooperate with supply chain partners.** Although the concepts of collaboration and cooperation among supply chain partners have

been around for a long time, achieving this has not been easy. The evidence presented in this study provides an economic rationale why supply chain partners must engage in these practices. The precursor for collaboration and cooperation is developing trust among supply chain partners, agreeing upfront on how to share the benefits, and showing a willingness to change from the old mindset. Once these elements are in place, supply chain partners must do joint decision making and problem solving, as well as share information about strategies, plans, and performance with each others. These activities can go a long way in reducing information distortion and lack of synchronization that currently plague supply chains and contribute to disruptions.

- **Invest in visibility.** To reduce the probability of disruptions, firms must be fully aware of what is happening in their supply chain. This includes internal operations, customers, suppliers, and location of inventory, capacity, and critical assets. The following may be needed to develop visibility:
  - Identify and select leading indicators of supply chain performance (suppliers, internal operations, and customers).
  - Collect and analyze data on these indicators.
  - Set benchmark levels for these indicators.
  - Monitor these indicators against the benchmark.
  - Communicate deviations from expected performance to managers at the appropriate levels on a real-time basis.
  - Develop and implement processes for dealing with deviations.
- **Build flexibility in the supply chain.** Firms must make careful and deliberate decisions to build flexibility at appropriate points in their supply chains to enhance responsiveness. There are multiple dimensions of flexibility and what will be appropriate for a firm depends on its operating environment.
  - **Building flexibility on the product design side.** Standardization, modularity, and use of common parts and platforms can offer the capability to react to sudden shift in demand and disruptions in delivery in parts.
  - **Building sourcing flexibility.** This can be achieved by using flexible contracts as well as use of spot markets to purchase parts and supplies. Spot markets can be used to both acquire parts to meet unexpected increase in demands as well as dispose of excess inventory if demand is below expectation.
  - **Building manufacturing flexibility.** This can be accomplished by acquiring flexible capacity that can be used to switch quickly among different products as the demand dictates. Firms should also consider segmenting their capacity into base and reactive capacity, where the base capacity is committed earlier to products whose demand can be accurately forecasted and reactive capacity is committed later for products where forecasting is inherently complex. Such would be the case for products with short

product life cycles as well as products with volatile demand. Late differentiation of products can also be used as a strategy to increase manufacturing flexibility.

- **Postponement strategy.** Postponement or delayed differentiation is a strategy that delays product differentiation at a point closer to the time when there is demand for the product. This involves designing and manufacturing standard or generic products that can be quickly and inexpensively configured and customized once actual customer demand is known. By postponing differentiation of products, the chances of producing products that the market may ultimately not want are minimized, thereby reducing the chances of demand-supply mismatches. Key success factors for implementing this strategy include:
  - Cross-functional teams that represent the design and manufacturing functions
  - Product and process reengineering to increase standardization
  - Modularity
  - Common parts and platforms
  - Collaboration with customers and suppliers
  - Performance measures and objectives that resolve conflicts and ensures accountability
- **Invest in technology.** Investment in appropriate technology can go a long way in reducing the chances of disruptions. Web-based technologies are now available that can link databases across supply chain partners to provide visibility of inventory, capacity, status of equipment, and orders across the extended supply chains. Supply chain event management systems have the ability to track critical events and when these events do not unfold as expected send out alerts and messages to notify appropriate managers to take corrective actions. This enables the firm to identify supply chain problems earlier rather than later and operate in a proactive rather than reactive mode. RFID technology has the promise to improve the accuracy of inventory counts as well as provide real-time information on the status of orders and shipments in transit and what is being purchased by customers. Such access to real-time information alleviates information distortions and provides true demand and supply signals, all of which can reduce the chances of demand-supply mismatches.

Although there are a number of strategies that firms can use to mitigate the chances of disruptions, which of these would be appropriate for a particular firm depends on the firm's operating environment. To identify what strategies to adopt, firms need a systematic process for risk management that is carefully and regularly applied. The process should be championed at the highest executive level as this is critical for bringing about awareness of the importance of managing disruption risk. A broad plan for developing and implementing such a process could be as follows:

1. **Assemble a cross-functional team of risk experts.** In most organizations, risk management is housed at the corporate level in insurance, legal and audit services. But supply chain disruption risks require a different type of arrangement. The knowledge of supply chain risks lies in marketing, operations, procurement, logistics, and information technology. Thus, the cross-functional team must include members from these areas as they have dealt in the past with disruptions and have sufficient experience to identify and quantify risks. To provide credibility and visibility to the team, top management must support and champion the team's activities and efforts by making a case for the importance of risk considerations and driving changes that mitigate risks.
2. **Characterize the major sources of risk.** The cross-functional team must regularly scan the internal and external environment to identify the vulnerable points of their supply chain. This involves identifying the primary sources of risk, estimating the probability of each risk happening, estimating the financial impact of the risk, the amount it would cost to recover from the risk, and the amount of time it would take to recover from the risk. Precise estimates on these issues may not be easy to get and therefore as a first step it would be appropriate to gather some qualitative data such as high or low frequency of occurring, high or low financial impact, and easy or hard to recover and so on.
3. **Assess and prioritize risks.** Once the primary sources of risk have been identified and agreed upon, the next step is to assess and prioritize the risks that should be of serious concern to the firm. Top management and the board should be made aware of the high risk issues. Various alternatives should be considered to mitigate the high risk factors. Such alternatives include developing contingency plans to deal with the risk should it surface, options for spreading risks through insurance, forward contracts, flexible contracts, and making organizational changes in how the supply chain is designed and operated so that these risks are mitigated in the future.
4. **Monitor risk and take actions as needed.** Once the primary risks issues have been identified and contingency plans have been developed, firms should set a system to monitor risks. Leading indicators need to be tracked, control limits need to be set to determine out of control conditions, two-way communication with suppliers and customers must be done on a continuous basis, and visibility systems must be in place. When risks surface the appropriate contingency plans are activated and the effectiveness of these plans in mitigating the risk is continuously monitored.
5. **Improve the risk management process.** Firms must continuously strive to improve their risk management processes. As and when risk is dealt with, efforts must be made to document the outcomes of the risk mitigation plans and highlight what worked and what did not work. These lessons should be shared across the organizations and used to improve the risk management process. Benchmarking a firm's process against other firms that have well-functioning risk management processes can identify best practices and help make a firm's process more robust and effective.

EXAMPLE 3.1 Lessons from Wal-Mart, Mattel, and Boeing Dealt on Dealing with Disruptions

The importance of managing supply chain disruptions raises the natural issue of what are some of the best strategies and practices in supply chain risk management. We highlight some of these using examples from Wal-Mart, Mattel, and Boeing.

How Wal-Mart responded to disruptions caused by Hurricane Katrina?

Hurricane Katrina, a low frequency event with a 1 in 200 annual odds, caused unimaginable devastation and disruptions in communities in Louisiana and Mississippi. Although the federal, state, and local governments failed miserably to respond to the devastation caused by Katrina, Wal-Mart was one of the few success stories in how organizations should respond to such disruptions. The key to Wal-Mart's success was a clear strategy for dealing with disruptions, detailed planning, and careful execution of the plan.

Wal-Mart started tracking and monitoring Katrina six days before the storm hit New Orleans. Using data from the National Weather Service and private meteorologists, Wal-Mart managers closely followed the storm's likely path and began shipping critical items to the distribution centers near the stores in the area where Katrina was likely to hit. These items were based on studies of customer buying patterns in hurricane-prone areas and the items that store managers usually need to ensure that their stores are operational. The trucking and transportation division was alerted to the need to load and ship critical items like backup generators and dry ice to stores at short notice. Backup communications plans with store managers and other key personnel were established and their roles and responsibilities in dealing with the disruption were reviewed and clarified. Plans were adjusted and modified on a real-time basis as Katrina changed its path. This detailed planning paid off as Wal-Mart turned out to be the only lifeline for many victims of Katrina. Wal-Mart provided relief days before FEMA could reach the affected areas, and was able to reopen it stores in record time, which provided further help and relief to its customers.

Wal-Mart success in dealing with Katrina highlights the importance of developing capabilities to deal with supply chain risks. Developing these capabilities requires leadership, commitment of resources, and detailed and meticulous planning. Building robust capabilities for dealing with supply chain risks involves the following steps:

1. **Analyze what could potentially go wrong.** This may require brainstorming, thinking about the unthinkable, observing disruptions that

your company and other companies have experienced, and involving experts in creating scenarios of what could go wrong.

2. **Identify and analyze possible alternatives to deal with different types of risks.** This may require benchmarking of best practices with other companies, scenario analysis, and idea generation. Various alternatives should be considered to mitigate the high-risk factors. Such alternatives include developing contingency plans to deal with the risk should it surface, options for sharing and transferring risks through insurance, forward contracts, flexible contracts, and making changes in how the supply chain is designed and operated so that these risks are mitigated in the future.
3. **Develop plans to deal with disruptions.** This involves outlining what needs to be done to deal with disruptions, when it will be done, how it will be done, and who will do it. The plan needs to assign responsibility and authority to employees to carry out the plans. Without such plans, employees are left clueless about what to do, which actually creates more chaos and magnifies the negative consequences of disruptions.
4. **Monitor the situation.** Companies should develop a system to monitor risks. Leading indicators need to be tracked, control limits need to be set to determine out of control conditions, two-way communication with suppliers and customers must be done on a continuous basis, and visibility systems must be in place.
5. **Execute the plan.** When disruptions occur, the appropriate plans are activated and the effectiveness of these plans in mitigating the negative impact is continuously monitored and adjustments need to be made on real-time basis.

EXAMPLE 3.2 **Mattel's product recall of lead-tainted toys**

Mattel Inc.'s recall (2007) of nearly one-million lead-tainted toys underscores the challenge companies face when they source globally in search of low costs. The toy industry has moved so much manufacturing to China to cut costs that now 80% of the toys that come to the U.S. are made in China. The relentless pressure that Chinese manufacturers face to cut costs creates incentives for manufacturers to cut corners to reduce costs. Lead paint is not only cheap and readily available but its use can speed up the production process, all of which leads to lower costs. In addition, Mattel allowed the manufacturer to do its own testing because it had a trusted 15-year relationship with the manufacturer but it did not perform the tests. Furthermore, the regulatory agencies did not have the resources to police the large volume of toy imports from China.

There are four key lessons for managing supply chain risks from Mattel's recall. First, relentless focus on cost reduction can often have unintended consequences. Companies should consider backing off somewhat on cost reductions to avoid creating incentives where suppliers cut corners on quality and safety. Second, even if you are not responsible for the disruption, you still pay. Interestingly, lead-tainted toys accounted for about 5 percent of the total toys recalled by Mattel. Yet, the damage to the overall reputation of Mattel's brand and image from the recall is far more than the direct cost of recalling 5 percent of the toys. Third, while much has been said about building long-term relationships with suppliers and trusting suppliers, companies must still be very watchful and monitor the processes at their key suppliers, particularly those that affect safety and health issues. Finally, as supply chains become more global, companies must make sure that they have traceability capabilities in their supply chains. This is critical because without isolating the source, it is difficult to solve the problem. This issue has become urgent because of contamination problems in food products, pet foods, and pharmaceuticals. The most recent case is the recall of the blood-thinning drug heparin. Nearly 80 deaths in the U.S. are attributed to contamination in heparin. The lack of traceability in the heparin supply chain has made it very difficult to trace the source of the problem to address the contamination issue.

Incidents such as the product recall by Mattel have clearly caused customers to become skeptical about products that are being sourced from China. This is a critical issue that affects all big retailers who depend on overseas supplier. To get better control of their supply chain and restore customer confidence, big and influential retailers like Wal-Mart, Target, Toys "R" Us, and Sears are requiring their suppliers to meet a new set of children-product safety requirements that goes far beyond existing government regulations. They are also encouraging suppliers to mark children's product with traceability information including the factory where the goods were made so that corrective actions can be taken should the products have quality and health issues.

EXAMPLE 3.3 **Boeing's Dreamliner delays**

The Boeing 787 Dreamliner has been very popular with orders for 892 planes from 60 airliners and delivery slots sold out beyond 2014. Unfortunately, Boeing cannot meet the promised delivery dates for these planes. Recently, Boeing announced the third delay in the production of the plane, pushing first deliveries at least 15 months later than initially promised.

The delay in the Dreamliner is an example of how outsourcing and globalization can create significant supply chain risks, which if not managed well can derail a company's best laid plans. To lower the cost to develop the plane on its own, Boeing outsourced the design and build of major sections of the aircraft to suppliers. The supply chain is quite global with nearly 15 major suppliers across nine countries. For example, the forward fuselage and the wing are manufactured in Japan, the center fuselage and the horizontal stabilizer in Italy, the wing tips in Korea, the trailing edge in Australia, the landing gear in the UK, the cargo access doors in Sweden, and the passenger-entry doors in France. Each of these first-tier suppliers uses its own set of suppliers, and so on, resulting in a highly complex and distributed supply chain. The subassemblies are transported by ship, air, road, and rail to facilities around Seattle for final assembly.

The supplier problems range from language barriers to glitches when some suppliers further outsourced the work. Boeing overestimated the ability of suppliers to do the tasks that Boeing could do with years of experience. Boeing also did not have deep insight of what was actually going on in the factories of the suppliers. Suppliers faced major issues in ramping up capacity. Coordinating across the various suppliers across the globe turned out to be more challenging than Boeing anticipated.

To deal with the delays Boeing has bought out the interest that one of its suppliers had in a joint venture. Boeing has hinted that similar moves are under consideration as it attempts to take control of key parts of its supply chain. Boeing managers are taking a more aggressive role in getting insight into suppliers' operations, including stationing Boeing employees in every major supplier's factory. Boeing is also trying to build real-time visibility of their suppliers' operations as well as developing a better understanding of how the plane comes together. Boeing's CEO is actively engaged in monitoring the plane's progress. Companies should be careful when they make outsourcing decisions and must balance the benefits of expected cost savings against the increased costs of managing supply chain risks. Boeing strategy underscores the limits and hazards of outsourcing.

3.6 Summary

The evidence presented in this chapter makes a compelling case that ignoring the risk of supply chain disruptions can have serious negative economic consequences. Based on a sample of more than 800 supply chain disruption announcements, the evidence indicates that firms that suffer supply chain disruptions experience 33% to 40% lower stock returns relative to their benchmarks, 13.5% increase in share price volatility, 107% drop in operating income, 7% lower sales growth, and 11% increase in costs. By any standard these are very significant economic loses. More importantly, firms do not quickly recover from these losses. The evidence indicates that firms continue to operate for at least two years at a lower performance level after experiencing disruptions. Given the significant economic losses, firms cannot afford such disruptions even if they occur infrequently.

The evidence presented in this chapter underscores why supply chain management issues deserve close attention by senior executives and board members. Heightened scrutiny of corporate governance makes executives more directly responsible for earnings forecasts and prediction. To the extent that supply chain disruptions can devastate corporate performance, senior executives must be fully aware of the performance of their supply chains.

As discussed, overemphasis on efficiency and removing slack from the system can make supply chains vulnerable, unreliable, and non-responsive. While efficient and lean supply chains are desirable objectives, they should not come at the expense of reliability and responsiveness. There is a trade-off between efficiency of supply chains and risk of disruptions within supply chains.

It is quite common to find practitioners and academics talk about changes in supply chain management practices and investments in terms of their effect on efficiency and cost savings. Risk issues are often ignored because they cannot be easily quantified. Yet the evidence presented in this chapter strongly suggests that investing in supply chain reliability and responsiveness is equally important, if not more, as investing in cost reduction. Such investments should be viewed as insurance against avoiding shareholder value destruction should disruptions happen. Given the evidence presented in this chapter, Senior management must ask the question of whether they can afford not to proactively prevent and manage supply chain disruptions risk.

APPENDIX 3.1: Methodology Used To Estimate Financial Impact of Supply Chain Disruptions

A. Methodology Used To Estimate Stock Price Performance

The basic idea in long-term stock price studies is to estimate abnormal returns for a sample of firms that have experienced the same kind of event, and then test the null hypothesis that the abnormal returns over the period of interest are equal to zero. An abnormal return is the difference between the return on a stock and the return on an appropriate benchmark, where the benchmark is chosen to control for factors that are known to explain normal stock returns. The abnormal return is the return that can be attributed to the event under consideration, and hence measures the effect of the event. The idea is that after controlling for the known factors, whatever remains unexplained is deemed as abnormal and can be attributed to the event under consideration. The current consensus seems to be that abnormal returns should be computed after controlling for size, market-to-book ratio, and prior performance.

Another critical issue with long-term stock price studies is the interpretation of the statistical significance of the observed long-run abnormal returns. Test statistics from many commonly used methods (such as comparing to S&P 500 Index) are severely mis-specified making it hard to judge the true significance of observed abnormal returns. Recent academic studies suggest that abnormal returns computed using matched portfolios or one-to-one matching give well-specified tests. Both these approaches are briefly described next.

ABNORMAL RETURNS USING MATCHED PORTFOLIOS

The matching portfolio approach computes abnormal returns using portfolios of firms that are similar in size, market-to-book ratio of equity, and prior performance as a benchmark. This approach is implemented using the following three-step procedure:

Step1 In each month, all eligible NYSE firms are sorted into deciles according to their market value of equity. Next all AMEX and NASDAQ firms are placed into the appropriate size portfolio. The smallest size decile portfolio is further divided into quintiles, resulting in 14 size portfolios. Each of the 14 portfolios is further divided into quintiles according to their market-to-book ratio of equity, resulting in 70 portfolios. The 70 portfolios are further divided into 3 portfolios each based on the stock price performance of firms in that portfolio over the previous year, resulting in 210 portfolios for each month where firms in each portfolio are similar in terms of size, market-to-book ratio, and prior performance.

Step2 In step 1, each sample firm has been assigned to a portfolio. The portfolio that a sample firm is assigned to 12 months before the month of the announcement date (the beginning of the measurement period) is identified. Since all other firms in this portfolio are similar to the sample firm on size, market-to-book ratio, and prior performance, all these firms can be considered as matched to the sample firm. The portfolio assignment and hence the set of matched firms for a sample firm remains the same over the three-year time period.

Step3 The buy-and-hold return for each sample firm is computed. If the sample firm is delisted before the end of a time period, the buy-and-hold return stops on the delisting date of the sample firm. The buy-and-hold return of each matched firm in the portfolio that the sample firm is assigned to is also computed over the same time period. If a matching firm is delisted prior to the end of the period or before the sample firm's delisting date, whichever is earlier, the overall stock market's value-weighted return is spliced into the calculation from the day after the matched firm's delisting date. This assures that the buy-and-hold return of the sample and matched firms are computed over the same time period. The benchmark return for each sample firm is then the average of the buy-and-hold returns of all its matched firms in its assigned portfolio. Abnormal performance is the difference between the return of the sample firm and the return to its assigned portfolio.

ABNORMAL RETURNS USING ONE-TO-ONE MATCH SAMPLES

In the one-to-one matching approach each sample firm is matched to an appropriately chosen control firm. The potential candidates for matching to a sample firm are those firms that belong to the portfolio that the sample firm is assigned to in the portfolio matching approach. This ensures that the matched firm will at least be similar to the sample firm on size, market-to-book ratio, and prior performance. Three different one-to-one control samples are created as follows:

1. Select the firm that is closest in size to the sample firm from the sample firm's assigned portfolio (size matched).
2. Select the firm that is closest in terms of prior performance to the sample firm from the sample firm's assigned portfolio (performance matched).
3. Select the firm that has the best matching on SIC code to the sample firm from the sample firm's assigned portfolio. If at least a one-digit match is not possible, the sample firm is dropped from the analysis (industry matched).

The abnormal return for a sample firm is the difference between its buy-and-hold return and that of the control firm.

B. Methodology Used To Estimate Changes in Share Price Volatility

The effect of supply chain disruptions on share price volatility is examined by comparing the standard deviations of stock returns before and after the disruption announcement date. Standard deviation of stock returns is estimated for four years, starting two years before through two years after the disruption announcement. Each year consists of approximately 250 trading days.

Many studies that consider the effect of corporate events on risk changes perform their analysis based on a comparison of the risk levels of sample firms before and after the event date. This approach could mis-estimate the true risk changes as risk can be influenced by certain macro factors that may have nothing to do with the event under consideration. Such factors could include interest rates, investor sentiments, consumer confidence, market expectations, global business environments, and so on. To control for such factors, the percent changes in the standard deviations of stock returns of the sample firms are compared against that of a matched control sample. In other words, the abnormal change in standard deviations of stock returns is estimated. For this purpose, the control samples used are the same ones used to estimate the buy-and-hold abnormal returns calculation using one-to-one matching approach (see Section A).

C. Methodology Used To Estimate Changes in Profitability

To provide a benchmark for the performance of the sample firms in the absence of disruptions, and to control for potential industry and/or economy wide effects, the performance of each firm in the sample is compared against an appropriately chosen control firm. It is reasonable to assume that firms in the same industry and of similar size are subject to similar economic and competitive factors. Thus, controls are chosen to be similar to the size and industry characteristics of the sample firms. To obtain the control-adjusted (or abnormal) change in performance, the difference between the change in performance of sample and control firms is estimated and tested for statistical significance.

The commonly used matching process in this type of analysis matches on size and SIC code. A composite measure of size is used in which sales and assets are each equally weighted.

To control for size and industry, three control samples are generated where each control sample is designed to address a specific potential bias or weakness in the others. In the first control sample, referred to as industry-size-matched control, each sample firm is matched to a control firm that has sufficient financial data available, has at least the same three-digit SIC code, and is closest in size, with the constraint that the ratio of size of the sample firm and control is always less than a factor of 3.

The second control sample, referred to as industry-matched control, attempts to find a control firm that has sufficient financial data available, has at least the same three-digit SIC code, and is closest in size. The key difference between this control sample and industry-size-matched control sample is that we do not put any constraint on the closeness of size matches.

The third control sample, referred to as most-matched, attempts to find a control firm that has at least the same two-digit SIC code and is closest in size. The key difference between this control sample and industry-matched control sample is that we allow for two-digit SIC code matches.

REFERENCES

Aviv Y. 2001. The effect of collaborative forecasting on supply chain performance. *Management Science*, **47**:1326–1343.

Barnes-Schuster D., Bassok Y., Anupindi R. 2002. Coordination and flexibility in supply contracts with options. *Manufacturing & Service Operations Management*, **4**:171–207.

Billington C., Johnson B., Triantis A. 2002. A real options perspective on supply chain management in high technology. *Journal of Applied Corporate Finance*, Winter **15**:32–43.

Cachon G., Fisher M. 2000. Supply chain inventory management and the value of shared information. *Management Science*, **46**:1032–1048.

Cachon G., Lariviere M.A. 2001. Contracting to assure supply: How to share demand forecasts in a supply chain. *Management Science*, **47**:629–646.

Casey N., Pasztor A. 2007. Safety agency, mattel clash over disclosures. *The Wall Street Journal*, September 4, A1.

Chopra S., Sodhi M.S. 2004. Managing risk to avoid supply-chain breakdown. *Sloan Management Review Fall* **46**:53–61.

Craighead C.W., Blackhurst J., Rungtusanatham M.J., Handfield R.B. 2007. The severity of supply chain disruptions: Design characteristics and mitigation capabilities. *Decision Sciences*, **38**:131–156.

Fairclough G. 2008. How a heparin maker in China tackles risks—Plant now supplying U.S. must make frequent checks to monitor difficult market. *The Wall Street Journal,* March 10, B1.

Ferrer J., Karlberg J., Hintlian J. 2007. Integration: The key to global success. *Supply Chain Management Review*, March 1–4.

Fisher M. 1997. What is the right supply chain for your product? *Harvard Business Review*, March–April, **75**:105–116.

Frohlich M. T., Westbrook R. 2001. Arcs of integration: an international study of supply chain strategies. *Journal of Operations Management*, **19**:185–200.

Hendricks K.B., Singhal. V.R. 2003. The effect of supply chain glitches on shareholder value. *Journal of Operations Management*, **21**:501–522.

Hendricks K.B., Singhal V.R. 2005a. An Empirical analysis of the effect of supply chain disruptions on long-run stock price performance and risk of the firm. *Production and Operations Management*, **14**:35–52.

Hendricks K.B., Singhal V.R. 2005b. Association between supply chain glitches and operating performance. *Management Science*, **51**:695–711.

Kesmodel D. 2008. Oversight Flaw Led to Meat Recall—Lax monitoring seen at plant's Cattle Pens Before Closure. *The Wall Street Journal,* 11 March, B1.

Kilgore M. 2003. Mitigating supply chain risks. White paper. Atlanta, Georgia: Chainalytics LLC.

Kleindorfer P.R., Saad. G.H. 2005. Managing disruption risks in supply chain. *Production and Operations Management Journal*, **14**:53–68.

Lakenan B., Boyd D., Frey E. 2001. Why Cisco fell: Outsourcing and its peril. www.strategy-business.com, Reprint no. 01306.

Lee H., Padamanabhan P., Whang S. 1997. Information distortion in supply chain: The bullwhip effect. *Management Science*, **43**:546–558.

Lunsford, J.L. 2007. Jet blues: boeing scrambles to repair problems with new plane—Layers of outsourcing slow 787 production; Hostage to Suppliers, *The Wall Street Journal*, December 7, A1.

Milner J.M., Kouvelis P. 2002. On the complementary value of accurate demand information and production and supplier flexibility. *Manufacturing & Service Operations Management*, **4**:99–113.

Narasimhan R., Jayaram. J. 1998. Casual linkages in supply chain management: An exploratory study of North American manufacturing firms. *Decision Sciences*, **29**:79–605.

Radjou N. 2002. Adapting to Supply Network Change. Cambridge, Massachusetts: Forrester Research Inc.

Rosenzweig E.D., Roth A., Dean J.W. 2003. The influence of an integration strategy on competitive capabilities and business performance: An exploratory study of consumer products manufacturer. *Journal of Operations Management*, **21**:437–456.

Sheffi Y. 2005. *The Resilient Enterprise: Overcoming Vulnerability for Competitive Advantage*. The MIT Press: Cambridge, Massachusetts.

Shin H., Collier D.A., Wilson D.D. 2000. Supply management orientation and supplier/buyer performance. *Journal of Operations Management*, **18**:317–333.

Smyrlis L. 2006. Risky business: There has been a significant increase in external risk during the past five years and supply chain is the major concern, new survey shows. *Canadian Transportation and Logistics* June 1, 14.

Taylor T.A. 2002. Supply chain coordination under channel rebates with sales effort effects. *Management Science*, **48**:992–1007.

Tang C. 2006. Robust strategies for mitigating supply chain disruptions. *International Journal of Logistics: Research and Applications*, **9**:33–45.

Tomlin B. 2006. On the value of mitigation and contingency strategies for managing supply chain disruptions risks. *Management Science*, **52**:639–657.

CHAPTER FOUR

Operational Strategies for Managing Supply Chain Disruption Risk

BRIAN TOMLIN AND YIMIN WANG

4.1 Introduction

Genzyme Corporation is a large biotechnology firm that develops and produces a range of pharmaceuticals targeted at a variety of medical conditions. On June 16 2009, Genzyme Corporation announced that it had discovered the virus Vesivirus 2117 in one of the bioreactors at its plant in Allston, Massachusetts.[1] While the virus strain is not thought to be harmful to humans, it does interfere with production efficiency. Genzyme made the decision to shut down production of the three drugs—Cerezyme, Fabrazyme and Myozyme—produced in the plant. In the fiscal year 2008, Cerezyme and Fabrazyme (used for the treatment of Gaucher and Fabry diseases—both genetic disorders) accounted for $1.7 billion of the company's $4.6 billion in revenue. Genzyme anticipated that the Allston plant would be back up

[1] All references to Genzyme in this chapter are based on information contained in press releases, communications and financial statements issued by the company.

Handbook of Integrated Risk Management in Global Supply Chains, First Edition. Edited by Panos Kouvelis, Lingxiu Dong, Onur Boyabatli, and Rong Li.

and running by the end of July. However, because of the long processing lead time associated with biopharmaceuticals, production launched in August would not yield product until later in the year. At the time of the disruption, Genzyme was in the late stages of constructing a second facility in Framingham, Massachusetts for the production of Cerezyme and Fabrazyme. While the Framingham plant would provide an added layer of protection against any future interruptions in the production of Cerezyme and Fabrazyme, Genzyme's only protection in June 2009 was its existing inventories of these two drugs. Unfortunately, the company's stockpile was not large enough to fully absorb the production loss. Genzyme's July 22 press release stated that

> Cerezyme and Fabrazyme inventories are not sufficient to avoid shortages during the period of suspended production and recovery. Genzyme is working closely with treating physicians, other health care providers, patient communities and regulatory officials worldwide to support patients with Gaucher and Fabry disease during the temporary period of supply constraint.

The press release estimated that the revenue loss associated with the disruption would be in the range of $250 to $600 million depending on the usability of the pre-disruption work-in-process inventory. Presumably the revenue loss would have been much greater if Genzyme did not have an inventory stockpile but would have been less if the Framingham plant had already come online. Fortunately for Genzyme (though not for the patients), there were no other existing Food and Drug Administration (FDA) approved drugs for the treatment of the Gaucher and Fabry diseases, and so the strategic risk of sustained market-share loss was not a preeminent concern in this instance. Genzyme did not anticipate shortages of Myozyme, the other drug produced in the Allston plant, in part because it had recently received approval from the European Commission for production of Myozyme at its facility in Geel, Belgium.

We can learn two important lessons from Genzyme's experience. The first lesson is that supply chain disruptions can and do have very large financial and strategic consequences. The second lesson is that operational strategies, if correctly designed and implemented, can effectively mitigate the financial and strategic risk associated with supply chain disruptions. It is the second lesson that is the focus of this chapter. Our primary purpose is to introduce supply chain practitioners and scholars to the operational strategies that firms can implement to manage supply chain disruption risk. By disruption we mean that a production facility or transportation link is (for some reason) temporarily unavailable or operating at a significantly lower capacity. Table 4.1 identifies and briefly describes five key operational strategies.[2]

[2] An operational strategy that we do not explicitly consider is the redesign of the product itself to reduce the complexity and riskiness of the supply chain required to support the production and distribution of the product.

TABLE 4.1 Five Operational Strategies for Managing Disruption Risk

Strategy	Implementation	Examples
Stockpile Inventory	Hold inventory that can be used to fill customer demand even if supply is interrupted.	In 2004, United Technologies Corporation temporarily increased its inventory buffer to protect against a potential supply disruption due to financial difficulties at a key supplier.
Diversify Supply	Source product from multiple vendors/facilities so that a problem at one vendor/facility doesn't affect the entire supply.	Nokia's multiple-supplier strategy reduced the impact of the 2000 Philips Semiconductor fire. Chiquita's multiple grower-location strategy reduced the impact of Hurricane Mitch in 1998.
Backup Supply	Have an emergency vendor, facility or logistics provider that is not normally used but that can be activated in the event of a supply problem.	Nokia responded to the Philips Semiconductor disruption by temporarily increasing production at alternative suppliers. New Balance responded to the 2002 U.S. West Coast dock disruption by rerouting ships to the East Coast and by airfreighting supplies.
Manage Demand	Influence demand to better match the actual supply by, for example, adjusting prices or offering incentives to encourage customers to purchase products that are less supply-constrained.	Dell responded to the disruption in memory supply caused by the 1999 Taiwanese earthquake by shifting customer demand to lower-memory computers.
Strengthen Supply Chain	Work with suppliers to reduce the frequency and/or severity of supply problems.	Unlike its competitor Xilinx, Altera does not source from multiple semiconductor foundries but works closely with foundry partner UMC to minimize yield-related supply problems.

In this chapter we explore each of these five strategies, with the goal of highlighting key factors that managers need to consider when designing and implementing their disruption mitigation strategy. In writing this chapter we made four important scope-related decisions that we wish to bring to the attention of the reader:

- While the lessons covered in this chapter are grounded in academic research, this chapter is not meant to serve as a review of the existing academic literature. It draws primarily, but not exclusively, on the research conducted by the authors. Where appropriate we will refer the reader to publications that underpin some of the recommendations offered in this chapter.

- While an effective risk management program comprises risk identification, assessment, response planning and ongoing monitoring and control, we do not attempt to address all four categories in this chapter. Our focus is on the operational strategies that firms can put in place to manage disruption risks and, as such, this chapter falls mostly in the response planning category. However, we will at times touch upon the other categories when they are relevant to understanding the advantage or disadvantage of a particular operational strategy.
- While supply chain management cuts across multiple firms, this chapter does not explore the incentives and informational issues that complicate multifirm disruption management. Nor does it focus on the demand-risk benefits of sourcing strategies. These topics are explored in Chapters 14 and 15, respectively.
- While some people might take a more expansive definition of supply chain disruption risk, this chapter will limit its definition to the risk of significant deviations between the delivered/produced quantity and the required quantity. As such, cost and quality risk will not be considered.

The rest of the chapter is organized as follows: In Sections 4.2 to 4.6 we discuss each of the five strategies described in Table 4.1. We then conclude in Section 4.7 by discussing the value of deploying multiple strategies and by identifying some directions for future research to advance the knowledge and practice of supply chain disruption management.

4.2 Stockpile Inventory

The concept of using inventory to protect against disruptions is a simple one. A company builds up a stockpile of inventory that can be used to fill demand during a disruption.[3] As compared to the other strategies listed in Table 4.1, it is a relatively easy strategy to implement, in part because it does not require coordination with suppliers or customers.

While Genzyme may well have been correct in choosing inventory as its disruption mitigation strategy, its experience during the Allston plant disruption offers a cautionary lesson. Genzyme could not immediately use its stockpile of Cerezyme inventory to meet patient demand during the virus-induced production disruption as the virus might also have contaminated the inventory. Before releasing this inventory, Genzyme had to prove to the FDA and the European Medicines Evaluation Agency that the inventory was not contaminated. Even after obtaining this approval, the inventory stockpile was not large enough to prevent

[3] While inventory is also useful for managing demand uncertainty, we focus here on the disruption-management motive.

shortages. The company had less inventory stockpiled than originally thought (because some of Allston capacity was being used for production of Myozyme) and the production disruption was longer than originally anticipated because Genzyme decided to conduct a more extensive sanitization of the Allston facility.

Although the inventory stockpile undoubtedly helped Genzyme navigate a difficult time, their experience suggests that inventory has its weaknesses as a disruption strategy. When evaluating inventory as a possible strategy, four important factors need to be considered: risk profile, detection, isolation, and recovery.

4.2.1 RISK PROFILE

In standard risk-management processes, risks are sometimes categorized along two dimensions: likelihood and severity. This categorization is also helpful in disruption management. Some disruptive events, machine breakdown for example, may occur relatively often (i.e., high likelihood) but the associated interruption may be short (i.e., low severity). Other disruptive events, natural disasters, may occur rarely (i.e., low likelihood) but the associated interruption may be very long (i.e., high severity).

This risk-profile distinction is particularly important when evaluating inventory as a strategy for mitigating disruptions. The severity (length) of a disruption determines how much inventory a company would need to fully protect itself against any supply interruptions. For disruptive events that are frequent but short, a company does not need to stockpile much inventory to protect itself. However, for events that are rare but long, a company would need to stockpile a very large quantity of inventory. This presents two problems—opportunity cost and temptation.

While the direct cost of storing the inventory (warehousing, labor, insurance, etc.) might be tolerable, there is a large opportunity cost associated with a large stockpile. The company has invested money in creating the inventory and it is not turning this money into revenue. This expense may not show up on an income statement but it is a substantial hidden cost of doing business and one that reduces a company's inventory turns and hence its return on assets.[4]

The second problem is one of temptation. For frequent-but-short disruptions, managers observe the inventory serving its intended purpose of buffering against production interruptions. For rare-but-long disruptions, managers feel the pain of lower inventory turns but might not observe the protection benefit because the disruptive event has not occurred. It is naturally tempting to drain the stockpile to boost inventory turns based on the assumption that a disruption won't occur soon because one has not occurred for a long time. For rare-but-long disruptions, the opportunity cost of inventory makes it an economically unattractive strategy and the temptation issue can render inventory an unsustainable strategy unless there are disciplined processes for maintaining the stockpile. In contrast, inventory can be a very effective strategy for protecting against frequent-but-short disruptions.

[4] For short lifecycle products, this problem is exacerbated because unsold inventory is obsolete at the end of the product lifecycle.

Of course, a company might opt for an inventory stockpile to protect against rare-but-long disruptions if its particular set of circumstances renders the other strategies listed in Table 4.1 even less attractive.

4.2.2 DETECTION

The fundamental problem with using inventory to protect against rare-but-long disruptions is that a large stockpile has to be carried for a long time. What if a company could adapt the size of its stockpile, increasing it when the threat of disruption was high and reducing it when the threat of disruption was low? By tailoring the stockpile size to the current level of risk, a company can alleviate the opportunity and temptation issues, thereby making inventory a much more attractive strategy. Adaptive strategies are becoming a reality, as evidenced by this description of United Technology Corporation's[5] use of supplier monitoring software:

> The software toolset uses pattern recognition technology to constantly monitor supplier data to determine if any of UTC's 18,000 suppliers are heading for trouble. In August 2004 the system generated a financial alert based on a recognized pattern of events for a key castings supplier. That partner was immediately identified as being important to a number of product lines, and a system-generated e-mail was sent to the OTL staff warning of a potential bankruptcy . . . [and] UTC increased its inventory buffer as an added layer of protection.[6]

Implementing an adaptive inventory strategy requires certain capabilities on the part of the company. First, the company needs some form of ongoing threat-detection process that monitors potential disruptive events and effectively detects and distinguishes between levels of risk. Therefore, an adaptive strategy is best suited to disruptive events in which (1) the risk evolves over time and (2) the firm can assess changes in the risk. Internal disruptions (such as labor stoppages) might fall within this category. Second, the company must have the ability to rapidly respond to an increase in the risk level. That is, it must have sufficient capacity to rapidly build up the stockpile when necessary. Otherwise, the protection level lags far behind the risk level and the adaptive strategy fails. In short, inventory becomes more attractive as the firm is better able to sense and respond to disruption risks.

4.2.3 ISOLATION

Inventory can help protect against the consequence of a disruption only if the inventory is usable and can be delivered to the demand location(s). If the underlying disruptive event damages the inventory or prevents its release, then the stockpile offers no protection. Fortunately for Genzyme and its patients, the virus did not

[5] United Technology Corporation's (UTC) is a large and diversified company that produces and sells complex, engineered products (e.g., elevators, air conditioners, and aircraft engines).
[6] Excerpt from Global Logistic & Supply Chain Strategies, December 2005.

contaminate the inventory stockpile. Phillips Semiconductor was less fortunate in 2000 when lightning caused a 10-minute fire in its Albuquerque, New Mexico plant. "Smoke particles had spread into the sterile room in the heart of the factory, contaminating the entire stock of millions of chips stored there" (Latour 2001).

If inventory is to be a company's chosen strategy, then it must strive to isolate the inventory from the disruptive events it is to protect against. If it is meant to protect against a hazard-induced plant failure, then the stockpile should not be stored in or near the plant. If it is meant to protect against a transportation-link failure, then it had better be stored on the customer side of the link. If the stockpile is not ring fenced but is continuously replenished in a first in–first out manner, then the company must be able to rapidly detect any contamination-induced disruption or else the stockpile will also be contaminated. In short, inventory is an effective safeguard only if it can be isolated from the disruptive event.

4.2.4 RECOVERY

The effect of a disruption does not end when the interrupted facility comes back online. Due to production lead times, there may be a further delay before the facility actually produces finished product. Even after the point when demand is being met, the firm's inventory stockpile (assuming it has chosen that strategy) will need to be replenished. Until such time as the stockpile is rebuilt, the company is operating at a reduced level of protection.

A disruption that occurs during this recovery time is especially problematic as it coincides with a temporarily diminished resiliency. The longer it takes to rebuild the inventory stockpile the longer the period of heightened exposure. The primary drivers of the post-disruption recovery time are the production lead time and capacity. The lead time determines how long until finished product starts to flow and the capacity determines how rapidly the stockpile can be rebuilt. Inventory is rebuilt only if production exceeds demand. Therefore, the closer the capacity is to demand, the longer the time to rebuild the stockpile. To compensate for this extended recovery time, a company needs to increase its initial stockpile quantity because, in effect, the stockpile has to protect against the possibility of disruptions during recovery. In short, the longer the recovery time, the more inventory is needed and the less attractive a strategy it becomes.

In summary, inventory is a simple strategy but has substantial hidden costs and dangers if the above factors are not carefully considered. That being said, inventory has its place in disruption management and should not be arbitrarily ignored. Even if inventory is not the primary strategy for managing disruptions, a company might want to consider holding a small stockpile as a secondary strategy. This buys the company some valuable time at the onset of a disruption if the primary strategy cannot be instantaneously activated. For those readers wanting to delve more deeply into the quantitative analysis of the inventory strategy, please see, for example, Song and Zipkin (1996), Tomlin (2006), Tomlin (2009b), Tomlin and Snyder (2007), and references therein.

4.3 Diversify Supply

Genzyme has manufacturing facilities in Belgium, Ireland, England, and the U.S. Some facilities produce active ingredients, others engage in bulk production of final product, while others carry out filling and finishing operations. Some facilities perform more than one of these steps. The Allston plant carries out bulk production of key genetic-disease targeted products, (e.g., Cerezyme, Fabrazyme and Myozyme). Synvisc, a biosurgery product, is produced in Ridgefield, New Jersey in the U.S. The recently expanded facility in Geel, Belgium can produce protein-based product and so complements the Allston plant's capability and capacity. In February 2009, the European Commission granted Genzyme approval for larger-scale production of Myozyme at the Geel facility. The value and limitations of Genzyme's partially diversified manufacturing network was in evidence during the virus-induced interruption to its Allston production. Because Genzyme had the ability and approval to produce Myozyme in two locations, the Allston interruption only partially disrupted Myozyme production. However, Cerezyme and Fabrazyme were fully disrupted by the Allston problem. Additionally, because only a subset of Genzyme's product portfolio was produced in Allston, an interruption to that plant did not affect all of it products.

The Genzyme experience captures the essence of the diversified supply strategy.[7] By splitting production (or sourcing) across multiple facilities (or suppliers), a company partially protects itself against disruptions because a problem at one site only interrupts a portion of the company's product flow. If the nondisrupted site can also ramp up production (i.e., provide additional emergency capacity), then it offers backup protection. We cover that disruption strategy in Section 4.4 on backup supply. Creating a diversified supply network is not without its challenges, and the following factors need to be carefully weighed when evaluating/implementing a diversified supply strategy for disruption management.

4.3.1 COST

At a very basic level, there is a cost associated with diversifying a supply network. There can be significant investment costs incurred each time a new facility is built or a new supplier is qualified. Operating several sites or suppliers multiplies the fixed costs associated with facility and supplier management. Variable costs associated with coordination increase. In an effort to protect itself against geographically located disruptions, a company might source from multiple countries. This increases the average cost of goods sold assuming different locations have different operating costs. The economies of scale gained by concentrating an activity in one location are diluted by diversification. While it is true that diversification

[7] There are other reasons beyond disruption protection for creating a diversified network. The discussion in this chapter is not intended to cover the wider range of benefits and concerns. Instead, it focuses on disruption-related considerations.

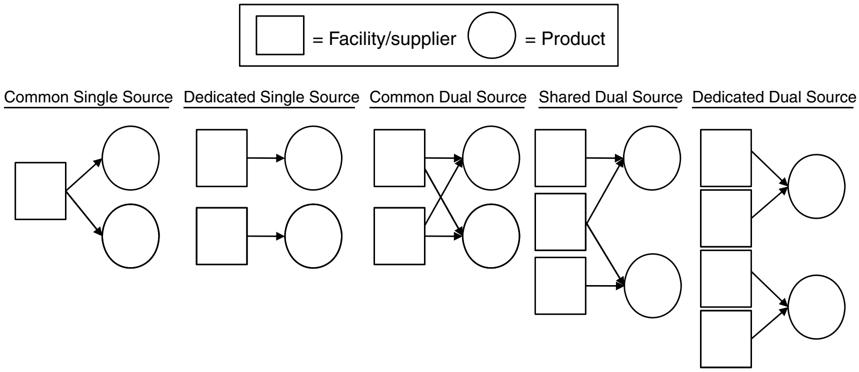


FIGURE 4.1 **Levels of supply diversification.**

can promote competition, which can drive down unit costs, the cost implications of a diversification strategy need to be carefully weighed against the benefits.

4.3.2 NETWORK CONFIGURATION

Diversification is not an all-or-nothing choice. Rather, the choice is the level of diversification the company desires to build into its product supply network. Using two products for ease of illustration, Figure 4.1 depicts network configurations that exhibit increasing levels of diversification (and cost) as one moves from left to right. At one extreme end of the spectrum, a company can concentrate production of all of its products in one facility (i.e., have no diversification protection). At the other end, a company can use multiple facilities for each product with no facility producing more than one of its products. This extreme form of diversification maximizes protection because a disruption to one facility only partially interrupts at most one product's supply. Figure 4.1 limits itself to dual sourcing but increased protection can be achieved by using more than two supply sites for a product.

Because of the cost and protection benefits associated with diversification, most companies will not want to position themselves at either end of the spectrum but instead will want to find an appropriate middle ground. The level of diversification and particular configuration chosen will depend not only on the company's assessment of the costs and protection benefits but also on technical, organizational, and regulatory constraints that the company operates within.

4.3.3 FAILURE CORRELATION

The protection offered by diversification rests upon the assumption that one facility continues to operate during a disruption to another facility. This assumption breaks down if a disruption interrupts both facilities. As the probability of both facilities simultaneously failing increases the value of diversification decreases. That is, the higher the correlation in failure across sites, the less protection diversification provides. Disruptive events arise from a number of underlying causes, some of which are specific to one site while others are (or at least can be) common across sites. For example, a strike might disrupt one supplier but a natural disaster might disrupt two suppliers located in the same region. The failure correlation increases with the number of common-cause events that pose a disruption risk. Put another way, a company can increase its diversification protection by reducing the likelihood of common-cause events.

Sourcing from multiple sites in the same geographical region introduces the possibility of a natural disaster disrupting all sites, and so sourcing from sites in different regions eliminates this common-cause disruption and increases diversification protection. Shared geography is not the only source of common-cause disruptions: Any common factor shared across sites introduces the possibility of common-cause disruptions. Genzyme believes that the Vesivirus 2117 contamination in its Allston plant originated in a raw material. Once Genzyme had developed an appropriate inspection technique, they were able to verify that the virus had contaminated production once already in the Geel plant. Fortunately that incident had already been resolved and so the virus did not induce a simultaneous interruption, but it could have. Along with geography and raw materials, managers need to investigate processes, systems, and policies when evaluating the potential for failure correlation across sites.

4.3.4 CONSISTENCY

By sourcing a product from multiple sites, a company introduces the potential for site-induced variations in their product supply. Depending on the company's tolerance for inconsistencies and the product's testability, consistency considerations can play an important role in the implementation of diversification. While one company might accept small deviations in product features, another company might require near-identical features such that products from two sites are fully interchangeable. For products, such as semiconductors, where the production process is a key driver of product variation, companies may need to replicate processes and equipments across sites if interchangeability of product is desired. Intel's well known "copy-exactly" production strategy is one such example. For products where identical functionality can be achieved using a variety of equipment or processes, there are more options available when diversifying supply.

The fitness-for-purpose of some products can be fully evaluated with a nondestructive test on the finished product. However, for other products, biopharmaceuticals for example, fitness-for-purpose can only be established by examination and verification of the production process. The FDA has stringent process validation standards for qualifying a new production process, as do equivalent agencies in other countries. A goal of this regulatory hurdle is to ensure consistent product quality. Before Genzyme could release production of Myozyme from its Geel facility, it had to gain approval from the European Commission because the Geel facility produced the product in larger bioreactors than used in the Allston plant. Those companies that require higher consistency and who are prone to process-induced variation, pharmaceutical companies being a prime example, face the most difficulty when implementing a diversification strategy.

In summary, diversification can be a powerful antidote to disruption risk but it is not without side effects. Diversification can increase costs and complexity. Also, network configuration decisions can be time consuming to implement and even more difficult to undo. Managers need to carefully weigh many factors when deciding how much diversification they need and how best to configure that diversification. A number of supply chain design software companies offer tools that address some of the trade-offs (e.g., fixed and variable transportation costs) inherent in network design, but no one tool captures all relevant factors. Managers therefore need to invest significant time and effort when crafting their diversification strategy. For those readers wanting to delve more deeply into the research underpinning some of the points discussed in this section, please see, for example, Dada et al. (2007), Tomlin (2009a), Tomlin and Wang (2005), and references therein.

4.4 Backup Supply

The inventory and supply diversification strategies require a company to make significant investments in advance of any potential disruption. The company must absorb the direct and indirect costs associated with the protection strategy whether a disruption occurs or not. A more appealing strategy might be one in which the cost is incurred only in the event of an actual disruption. One way to align the expenditure of resources with the occurrence of a disruption is to rely on a backup supply strategy. In this strategy, a source or site that does not routinely produce the disrupted product temporarily steps in to meet the supply requirement during the disruption to the primary source. Analogously, a company might avail of an emergency transportation mode if their standard mode is disrupted. For example, in response to the air-traffic disruption resulting from the September 11th terrorist attack, Chrysler temporarily shipped components by ground from the U.S. to their Dodge Ram assembly plant in Mexico.

If a company has a diversified supply strategy, then one of the sources that currently manufactures the product might be able to temporarily increase production. This would be a combined diversified/backup supply strategy. For example, when Genzyme's Framingham production facility comes online, it will have two approved sites capable of producing Cerezyme and Fabrazyme.[8] In addition to diversifying supply as a protection against future disruptions, Genzyme will have the option of temporarily increasing production at one facility during a disruption to the other facility. This will provide an added layer of protection. Incurring cost only when necessary is an attractive proposition but there are additional factors that need to be considered when evaluating the backup strategy.

[8] Mechanical construction of the Framingham plant is scheduled for late 2009 but due to the lead time associated with qualification runs and regulatory approval, Genzyme does not anticipate commercial production until 2011 for Fabrazyme and 2012 for Cerezyme.

4.4.1 AVAILABILITY AND COST

Just as the diversified strategy breaks down if all sources fail simultaneously, the backup strategy breaks down if the backup source is not available when called upon. If a company has excess capacity (capable of performing the disrupted activity) within its own network of facilities, then availability is not an issue.[9] However, if a company has to look outside its own network for emergency capacity, then availability may become a significant issue. First, there must be a third-party provider capable of performing the disrupted activity and this party must have (or be able to obtain) additional capacity. The more specialized the activity the fewer competent providers there are likely to be. Second, the company must be able to access the backup source's capacity.

Accessing backup capacity can be especially problematic if a disruption affects a supplier shared by two competitors. In the aftermath of the Phillips Semiconductor disruption, a key customer, Nokia, obtained emergency capacity from other Phillips Semiconductor facilities. Ericsson, a competitor of Nokia and also a customer of the disrupted plant, was not able to obtain additional capacity from Phillips as it had already been allocated to Nokia (Latour 2001). Also, Ericsson had rationalized its supply base and therefore had fewer qualified suppliers to source from during the emergency. Smaller companies may find that their ability to access scarce capacity is limited as suppliers allocate their capacity to more important customers.

Disruptions can temporarily create competition between companies that don't even operate in the same industry. The fight is not over customers but over access to a scarce resource. When multiple companies are interrupted by a common disruption, such as the 2002 West Coast port disruption in the U.S., they compete over access to whatever spare capacity exists. There was intense competition for airfreight capacity from Asia during the port disruption as firms attempted to circumvent the disruption by flying goods to the U.S. As demand outstripped supply, airfreight costs increased by 30% within a week and not all companies were able to access airfreight capacity even if they were willing to pay the price. Companies that had long-established relationships with the third-party transportation providers were (understandably) given preferential allocation.

To guarantee availability, a company may enter into a contract with a vendor to provide capacity in the event of an emergency.[10] Such a contract will normally stipulate how much capacity the vendor will provide and associated payments. The vendor needs to be compensated for providing emergency capacity, and this compensation might take the form of an upfront (or ongoing) payment

[9] Of course, just as with the diversified supply strategy, if a common cause event interrupts both the regular and the emergency source, then the backup strategy offers no protection. See the earlier discussion on correlation in the diversified strategy.

[10] Care must be taken when validating the provider's ability. If the provider is making commitments to multiple customers, does it have the capacity to simultaneously meet all of its commitments? If not, then at least one customer is going to be severely disappointed if a disruption simultaneously affects all the vendor's customers.

for reserved capacity and/or an agreement to pay a premium when the capacity is accessed. Either way there is often a higher variable cost associated with emergency processing. Even if backup production occurs within a company's network, there may be costs associated with overtime or expediting supplies to the emergency facility. A careful accounting of the associated costs needs to be undertaken when evaluating the backup strategy.

4.4.2 RESPONSE TIME AND MAGNITUDE

Assuming that the company can access backup capacity, the next question is how long does it take and how much backup capacity can it obtain? The shorter the response time (how long) and the higher the response magnitude (how much) the more protection is provided by the backup strategy. Response time can be broken into three parts: detection time, coordination time, and ramp time. Detection time refers to the time between the onset of the disruption and the acknowledgment by the company that it has a problem. In the Phillips Semiconductor disruption, Nokia detected the problem almost instantly while Ericsson did not react for several weeks.

To minimize detection time, companies must be vigilant in monitoring all supply and production activities for any hint of problems to come. Coordination time refers to the time between detection and agreement by a backup source to provide capacity. Here, prior planning and relationships are vital. Companies that engage in effective business continuity planning will have plans and assigned responsibilities for reacting to a disruption. If backup supply is the company's strategy, then continuity planning should have laid out the necessary steps involved in coordinating backup suppliers. Coordination will be faster if the company has an existing arrangement with the supplier for the provision of backup capacity. Prior qualification of the supplier (if possible) eliminates the time-consuming step of validating the backup source at the start of a disruption. Even after coordinating with the backup provider, there will be a delay in bringing the additional capacity online unless the provider has all the necessary supplies and idle labor and equipment of the required type. This ramp time (i.e., the time between agreeing to provide capacity and producing the extra units at the agreed volume) is the third element of the response time.[11]

Response magnitude refers to the amount of additional capacity that can be made available. This will be a function of the spare capacity (or readily accessible capacity) of backup sources and the number of sources. Those companies requiring highly specialized labor or equipment may find it difficult to find much additional capacity unless they have taken the (possibly prohibitively) expensive step of paying to ensure specialized assets are available when called upon. Companies can reduce response time and increase response magnitude through advanced planning by investing in additional internal capacity and by paying for

[11] The ramp time will be lower if the backup source is already producing the product, and so a company that implements a diversified supply strategy has an advantage in implementing a backup strategy because diversification complements the backup strategy.

preferential access to third party capacity. Assuming a limited budget, should a company expend more resources on improving response time or magnitude? It depends. In particular, it depends on the nature of the disruptions the company faces. For frequent-but-short disruptions, a backup strategy is essentially useless unless the response time is very short. As such, the firm should focus on reducing response time. It can store inventory to make up for the capacity shortfall caused by the response magnitude being insufficient to cover lost production.[12] For rare-but-long disruptions, response magnitude is crucial assuming that the response time is not egregious. The firm can store inventory to cover the lost production during the response time whereas using inventory to make up for a large ongoing capacity shortfall during an extended disruption would be very expensive.

In summary, the backup supply strategy is a very attractive strategy for protecting against rare-but-long disruptions because it aligns the protection expense with the reality of a disruption. The company (in theory) is not paying to protect against a hypothetical event but instead reacting to an actual event. However, companies can be overly optimistic about the effectiveness of their backup strategy. When evaluating and/or implementing the backup supply strategy, managers need to engage in a realistic appraisal of availability, cost, response time, and magnitude. By engaging in scenario planning around these four dimensions, companies can better assess and implement a robust backup strategy. For those readers wanting to delve more deeply into the research underpinning some of the points discussed in this section, please see, for example, Chopra et al. (2007), Tomlin (2006), Tomlin (2009a), and references therein.

4.5 Manage Demand

In the three strategies discussed so far, company efforts are directed at managing the supply side of the disruption-related supply-demand imbalance. An alternative, or complementary, strategy is to manage the demand side of the imbalance. Genzyme devoted lots of energy to working with patients, physicians, health organizations, and regulatory authorities to mitigate the disruption effects.

Because the inventories of Cerezyme and Fabrazyme were insufficient to cover demand, Genzyme had to determine how to ration the drugs to minimize adverse health reactions. In collaboration with physicians and regulatory authorities, the company developed treatment protocols that altered the consumption rate for patients in an effort to match demand with available supply, with a goal of protecting the most at-risk patients. In addition, Genzyme made attempts to enable patients to switch to alternative treatments. Although there were no approved commercially available alternative to Cerezyme, Genzyme was in the clinical-trials stage for an investigational drug—GENZ-112638. Certain patients would be eligible to enroll in these trials. Also, Genzyme petitioned the FDA to allow the temporary prescription of GENZ-112638 even though it had yet to be approved. The FDA can grant approval for the use of investigational drugs if circumstances dictate it

[12] A company might want to rely solely on the inventory strategy for frequent-but-short disruptions.

is in the best interest of the patient population. The FDA also worked with Shire and Protalix, companies that were developing drugs to compete with Cerezyme, to consider applications for special approval of their investigational drugs as a means of generating additional supply.

As part of their broader disruption-recovery strategy, Genzyme actively communicated with patients, physicians, and other stakeholders as they crafted their rationing and switching plans. A dedicated website—http://supplyupdate.genzyme.com/—was created to keep patients informed of the ongoing status of the drug supply. Genzyme's experiences highlight crucial foundations of an effective demand management strategy—switching, rationing, and communication.

4.5.1 SWITCHING

If a company sells more than one product and customers exhibit some willingness to switch between products, then the company might be able to mitigate a supply disruption by inducing customers to switch from the supply-constrained product to a non-constrained product. In September 1999, an earthquake in Taiwan disrupted production of crucial supplies, including memory, used in personal computers. Dell Inc. reacted to the disruption by shifting customer demand to lower-memory computers.

Switching is an option only if the company has a supply of an alternative, acceptable product that it can make available to customers affected by the disruption. If the company's supply/production network is configured so that all cross-substitutable products flow through the same facilities, then all these products may be constrained by a disruption and demand switching would not be feasible. Our earlier discussion of supply diversification highlighted network configuration as a key lever. In Figure 4.1, certain configurations have complete overlap in the sources used for each product whereas other configurations exhibit some degree of product-source diversification, that is, production resources are not shared by all products. Companies wanting to use demand switching to manage supply disruptions need to configure their network so that it exhibits some level of product-source diversification.

Demand switching is not limited to managing supply-demand imbalances induced by disruptions. Dell Inc. has excellent demand switching capabilities that helps it manage uncertainties in demand. By judicious use of special offers and the option recommendations on its website, Dell is able to shape short-term demand to better match its incoming supply. Dell can do this because (1) it has developed excellent supply-visibility capabilities, that is, it continuously monitors the status of incoming supplies and inventory levels; (2) it has developed insight into how customer purchasing behavior responds to pricing and recommendations; and (3) it has an effective direct-to-customer website channel and so can directly influence its customers rather than relying on channel partners who might be just as happy to switch customers to a competitor's product. It was precisely because Dell had already developed these capabilities, that is was able to effectively respond to the Taiwanese earthquake disruption.

4.5.2 RATIONING

If a company cannot switch (enough) customers to balance supply and demand during the disruption-induced supply constraint, then it must decide how to allocate its limited supply amongst its customers. The appropriate rationing mechanism will depend on the goals of the company and the options its customers have to take their business elsewhere. Is there a key customer that the company must retain at all costs? If so, then filling this customer's demand first might be appropriate even at the cost of shorting all other customers. Is it important to treat, and to be seen to treat, all customers equally? If so, then a rationing rule that allocates customers a common percentage of their request might be appropriate, or, to avoid customer gaming, the allocation might be based on recent volumes purchased. If certain customers can easily take their future business elsewhere then the company might want to give a preferential allocation to these customers. None of these choices are easy as some customers will not be served, and there is a cost to not satisfying customers.

In some cases, the Genzyme case being a prime example, the customer cost should not be measured in monetary terms as rationing supply has the potential to harm a person's health. Supply rationing in these instances is a much more complex challenge. In the words of Geoff McDonough, a Genzyme senior vice president, the philosophy underpinning Genzyme's rationing program was to "preserve inventory for the most vulnerable patients and to ensure global equity in this extremely challenging time for patients and physicians." Genzyme's rationing program is best described by quoting from their August 10, 2009 press release:

> In the United States, Genzyme last week implemented a dose conservation program to try to ensure that the most vulnerable patients continue to receive Cerezyme. The company is now shipping Cerezyme only to two patient populations: patients with Gaucher disease type 1 who are 18 years of age or younger, and patients with Gaucher disease types 2 and 3. As part of U.S. dose conservation, Genzyme has also created an emergency access program, through which physicians may apply to receive Cerezyme for patients who are in life-threatening situations. Applications will be reviewed using criteria formulated in consultation with stakeholders from the physician and patient communities, and decisions will be made by a Genzyme medical committee with guidance from an independent group of physicians and patient representatives. Patient access via this program will be determined by available inventories going forward. Genzyme expects the U.S. dose conservation measures to remain in place until supply begins to normalize at the end of this year. This dose conservation program depends on the release of the two remaining finished Cerezyme lots but no work in process material. Outside of the United States, Genzyme is currently in discussions with regulatory authorities, physicians, and patient organizations to determine how to manage the supply of Cerezyme, and the company will begin shipping according to the revised inventory levels this week.

Genzyme's actions highlight the immense challenges that can arise in rationing. In theory, pricing offers an alternative to rationing. If supply is constrained, prices can be temporarily increased to reduce demand. This can be a dangerous strategy, however, if customers (or the general public) believe that the company is profiting unfairly from a problem of their own making. In industries where goods are bought and sold on a spot market, pricing may be a feasible lever for balancing supply and demand.

4.5.3 COMMUNICATION

Genzyme was very proactive in communicating its supply problem and rationing/recovery plans with those affected. This helped alleviate, if not eliminate, the concerns of patients. During the Taiwanese earthquake disruption, Dell induced customers to switch their purchases to lower-memory computers. Apple, on the other hand, did not possess Dell's demand switching capabilities but attempted to meet customer requests by shipping different product from what they ordered (Griffy-Brown 2003). Customers were understandably unhappy and Apple fared worse than Dell during the disruption (Sheffi 2005).

Proactive communication is crucial when implementing rationing or switching, as otherwise the company runs the substantial risk of damaging customer relationships and its wider brand equity as a result of unflattering news reports. An effective strategy needs to consider the audience (who), the message (what), the timing (when), and the medium (how) when developing its plans to manage customer and stakeholder communications. Executives would be well advised to immediately enlist the help of communication specialists in developing their communication plans, and employees should be informed about how to handle customer and press inquiries.

In summary, demand management can be an effective strategy for mitigating supply disruptions if the company has the necessary capabilities already in place. The supply-chain and customer-management processes and systems required for the switching element of the strategy can be difficult to develop and implement. It is more likely that a company would develop the switching capability to manage ongoing demand volatility than to develop it with the primary purpose of managing the risk of supply disruptions. However, if a company has the capability it should certainly avail of it during a disruption.[13] Rationing may be forced upon the company whether it has planned for this eventuality or not. Advanced planning is highly advisable to enable smoother implementation during the disruption. When evaluating the demand management strategy, managers need to be realistic about their own capabilities and the response of customers to switching incentives and rationing plans. For those readers wanting to delve more deeply

[13] Companies operating in industries with high demand volatility and high inventory-related costs are more likely to make the necessary demand-switching investments as they continually battle supply-demand imbalances.

into the research underpinning some of the points discussed in this section, please see, for example, Alizamir (1981), Tomlin (2009a), and references therein.

4.6 Strengthen supply chain

The four strategies covered so far all tackle the impact of a disruption, that is, they seek to minimize the negative consequences of a supply interruption. Companies can, and should, also consider addressing the likelihood of a disruption. By stress testing their operations and conducting effective scenario planning, companies can identify and rectify weakness in their current operations that leave themselves vulnerable to internal disruptions. Building strong internal processes does not go far enough. According to Genzyme's June 25 press release:

> The virus [that caused the Allston plant disruption] was likely introduced through a raw material used in the manufacturing process, and the company is collaborating with its suppliers to address this issue and implement steps to protect against recurrence. Genzyme is also evaluating adding steps to its raw-materials screening and virus-removal processes to make them more robust, including testing all of its raw materials for the presence of Vesivirus 2117 using the highly specific assay it developed. In addition, Genzyme is collaborating with other biologics manufacturers to learn from their experience and apply this knowledge to resolve the current situation and implement enhanced safeguards. Genzyme intends to share its own experience with this virus through appropriate mechanisms so that others within the industry may benefit. This includes working to ensure that an assay for Vesivirus 2117 becomes widely available to the industry.

As is common, Genzyme relied on suppliers to produce important ingredients. This meant that Genzyme was at risk of a supplier-induced disruption, either through incoming materials causing problems at a Genzyme facility or through a supplier disruption interrupting material flow to a Genzyme facility. Most companies find themselves in a similar position. Therefore, they must determine how best to protect themselves from supplier-induced disruptions. Companies can avail of the four strategies discussed so far but they can also develop their supply base to reduce the likelihood of supplier-related interruptions. When contemplating this strategy, companies need to consider their supplier-development approach; the timing of supplier commitment, and the risks of spillover.

4.6.1 SUPPLIER-DEVELOPMENT APPROACH

Framing supplier development in the wider context of increasing supplier "performance and/or capabilities to meet the firm's short- and/or long-term supply needs," Krause (1997) offers the following categorization of supplier development approaches: (1) enforced competition through sourcing from multiple suppliers; (2) incentives (i.e., the promise to a supplier of benefits such as increased volume), and (3) direct involvement, whereby the company exerts effort to improve its suppliers capabilities. Direct involvement can range from relatively low-effort activities such as informal evaluation and feedback, through medium-effort activities such as certification programs, to more effort-intensive activities such as supplier training programs and equipment investments.

According to their studies, (Krause et al. (2007)) found that delivery-reliability related improvement outcomes depended more on direct involvement, than did cost improvement outcomes. This suggests that enforced competition and incentives, while effective at reducing costs, may be less effective than the direct involvement approach at strengthening supplier reliability. Companies would therefore be well advised to work directly with suppliers on improving their reliability. In fact, a 2008 survey (Global Supply Chain Trends 2008–2010) by the consulting company PRTM (Cohen et al. 2008) finds that many companies are indeed taking this path: "companies have developed numerous ways to minimize disruption related to quality and delivery issues. Increasing the frequency of on-site audits is the most commonly cited approach, followed by physical deployment of their company's resources within the supplier's location, increased inspection, and increased supplier training. Other risk mitigation strategies mentioned frequently include consistent dual sourcing strategies." Well-known companies in a wide range of industries, including Intel in the electronics industry, Honda and Toyota in the automotive industry, and Kimberly Clark in the consumer goods industry, are engaged in direct-involvement supplier development.

4.6.2 COMMITMENT TIMING

Supplier improvement efforts can and do fail to achieve their desired outcome. A company attempting to improve a supplier's reliability might find that the supplier's processes are no more resilient despite the effort invested by the company. This poses something of a quandary for the company: Should it commit to the supplier (i.e., enter a binding agreement to use the supplier) before it observes the outcome of its improvement efforts or should it postpone commitment until improvement outcomes are known? Early commitment (i.e., before supplier improvement outcomes are known) assures the supplier that it will receive orders from the company and so provides an incentive for the supplier to engage in meaningful collaboration. This can be very beneficial if improvement outcomes depend on good-faith efforts on the part of both companies. Late commitment (i.e., after supplier improvement outcomes are known) enables the company to hedge against improvement failure because it gives the company the option of allocating its orders amongst suppliers after observing improvement outcomes.

Both timing tactics have merit, and some companies choose early commitment while others choose late commitment. Late commitment is less valuable if a company has a very heterogeneous supply base, that is, when potential suppliers (of the same component) differ significantly in cost or some other relevant dimension. Postponing supplier selection through late commitment offers little value as it is already obvious which supplier will be preferred regardless of improvement efforts. Late commitment is of particular value when improvement efforts at different suppliers exhibit different outcomes. If the improvement success probability is very high or very low, then it is likely that all improvement efforts will either succeed or fail, and so outcomes will be similar. Late commitment is therefore more valuable when improvement success probabilities are not too high or too low. If improvement costs are low, then the company can afford to engage in improvement efforts with multiple companies before allocating its orders. Low improvement costs therefore favor late commitment. Companies undertaking supplier development efforts need to carefully weigh the advantages and disadvantages of the commitment timing options.

4.6.3 SPILLOVER RISK

Companies often source from the same supplier as their competitors. By working with the supplier to improve its operations, the company runs the risk of unwittingly benefiting its competitor. The supplier may take the knowledge or assets gained during the improvement effort and improve its service offering to the competitor. In essence, one company can be a free-rider that gains from another company's supplier development efforts. A greater risk is that the supplier will unwittingly (or knowingly) pass along confidential information learned during the collaboration to the competitor. Companies need to weigh these risks and determine if safeguards can be put in place. On a more positive note, competitors might benefit from collaborating to address a common threat. Collaboration can spread the costs amongst more companies and might lead to a better solution.

In summary, companies can reduce the likelihood of supplier-induced disruptions by strengthening their supply chains through supplier development efforts. Oftentimes, supplier development programs grow out of one-time projects instigated in reaction to a particular supplier failure. Effective supplier development programs requires careful planning and implementation. In addition to the issues discussed above, managers need to determine which suppliers to target, how to allocate effort across their supply base, and how to measure outcomes. For those readers wanting to delve more deeply into the research underpinning some of the points discussed in this section, please see, for example, Krause (1997), Krause et al. (2007), Wang et al. (2010), and references therein.

4.7 Conclusions

Companies should carefully consider all five operational strategies—inventory, supply diversification, backup supply, demand management, and supply-chain strengthening—when developing their disruption-risk management plans. Each strategy has its strengths and limitations and managers need to align the strategy with the environment they operate in. A one-size fits all approach of employing the same strategy for all product lines may not be appropriate if different products exhibit different supply chain and market characteristics.

The disruption management strategy should be tailored to the needs of each product. Managers may also want to deploy multiple strategies (e.g., combine inventory with backup supply) to add an extra layer of protection if they are especially concerned about minimizing interruptions. Because some of these five strategies—inventory, backup supply, and demand management—help mitigate demand uncertainty, managers should not segregate supplyrisk and demand-risk planning. Instead, these five strategies should be viewed through the broader lens of supply-chain risk management, which encompasses availability, cost, demand, and quality risks.

The field of disruption risk management is still in its relative infancy, and many questions remain unanswered. Research and development that addresses the following needs would help advance the knowledge and practice of supply chain disruption management.

4.7.1 FROM INSIGHT TO DECISION SUPPORT

As one might expect in a nascent field, much of the scholarship to date on disruption risk management has focused on improving our understanding of the underlying phenomena. That is, research has helped shed light on the factors that need to be weighed when developing a disruption management strategy. Much of the quantitative work has been done in a "controlled setting" using simplified models that are amenable to analysis.

Moving forward, there is a need to develop decision-support tools that can help managers quantitatively evaluate various strategies in a realistic supply chain setting. Such a development would echo the evolution of network design and inventory-target setting from small-scale models with limiting assumptions to commercial-strength software applications. This will require models that are scalable to allow for hundreds, if not thousands, of stock keeping units (products) and hundreds of supply chain processing locations and links. A first step would be the development of large-scale models that enable performance evaluation of a chosen strategy. Going further, companies would benefit from large-scale models that also recommend appropriate strategies based on the manager's objective and budget. Models would also need to be sufficiently flexible to capture the needs of a variety of industries.

4.7.2 RISK EVALUATION

Much of the existing work on disruption risk management assumes that managers have some estimates of the likelihood and severity of disruptions. This may be reasonable in some circumstances but not in all. For disruptions caused by natural disasters or supplier bankruptcies, there may be publically available hazard or financial information that is helpful in estimating the underlying likelihoods. For recurring disruptive events, firms can refine their estimates based on past history. Companies would benefit from the development of robust methodologies and systems that enable firms to update their risk estimates based on past experience and public sources of information. However, even with such systems, it is difficult to estimate the probability of events that occur very rarely. If an event has not occurred in the past, that does not mean it cannot occur in the future, and so basing estimates on the past frequency of occurrence can be misleading.

As with any quantitative modeling approach, managers should engage in sensitivity testing to ensure that their plans are robust to misestimation of the disruption parameters. However, the disruption-management field would benefit from the development of appropriate methodologies that explicitly account for the difficulties in estimating and evaluating low-probability events. For an investigation of how people interpret and assess low-probability events in the context of accidents and disasters, we refer the reader to Kunreuther (2001).

4.7.3 INTERRUPTION INSURANCE

Companies can purchase special insurance policies that provide some coverage for costs or lost income associated with a business interruption. Typically the insurance policy is aimed at insurable events (e.g., fire-related damage) covered by a standard insurance policy. The standard policy helps to defray the cost of repairing/replacing the damaged facility, whereas interruption insurance targets the losses (including income) incurred as a result of the damaged facility interrupting normal business activity. Companies can even purchase coverage for interruptions caused by disruptions at a supplier's facility, assuming the disruption is caused by an insurable event.

Insurance, therefore, offers another tool for managing disruption risk. Under what circumstances is interruption insurance a better solution than an operational strategy? Is interruption insurance a substitute for an operational strategy or does it complement the operational strategies? There is a need for research to answer these and other questions to help bridge the disciplines of insurance and operations. We note that there is research that explores other non-operational mechanisms (e.g., contracts and financial instruments) used to manage disruption risk, see Babich et al. (2010) and Wadecki et al. (2010) for examples.

Thanks to the work of many scholars and practitioners, a solid foundation is being built for the field of disruption risk management. A good start has been made but much remains to be done to deepen our understanding and to develop solutions. Collaboration between scholars and practitioners is crucial to help steer the development of the field so that it addresses the needs and realities of business in a manner that is grounded in rigorous and scientific methodologies.

REFERENCES

Alizamir S., Alptekinoglu A., Sapra A. 1981. Demand management using responsive pricing and product variety in the presence of supply chain disruptions. Working paper, SMU Cox School of Business.

Babich V., Aydin G., Brunet P.-Y., Keppo J., Saigal R. 2010. Risk, financing and the optimal number of suppliers. In H. Gurnani, A. Mehrotra, and S. Ray, editors, *Managing Supply Disruptions*. London: Springer-Verlag.

Chopra S., Reinhardt G., Mohan U. 2007. The importance of decoupling recurrent and disruption risks in a supply chain. *Naval Research Logistics*, **54**:544–555.

Cohen S., Geissbauer R., Bhandari A., D'heur M. 2008. Global Supply Chain Trends 2008–2010. Sith Annual Survey by PRTM Management Consultants.

Dada M., Petruzzi N. C., and Schwarz L. B. 2007. A newsvendor's procurement problem when suppliers are unreliable. *Manufacturing & Service Operations Management*, **9**:9–32.

Griffy-Brown C. 2003. Just-in-time to just-in-case. *Graziadio Business Report, Pepperdine University*, 6.

Krause D.R. 1997. Supplier development: Current practices and outcomes. *Journal of Supply Chain Management*, **33**:12–19.

Krause D.R., Handfield R.B., Tyler B.B. 2007. The relationships between supplier development, commitment, social capital accumulation and performance improvement. *Journal of Operations Management*, **25**:528–545.

Kunreuther H., Novemsky N., Kahneman D. 2001. Making low probabilities useful. *Journal of Risk and Uncertainty*, **23**:103–120.

Latour A. 2001. Trial by fire: A blaze in albuquerque sets off major crisis for cell-phone giants. *The Wall Street Journal*.

Sheffi Y. 2005. *The Resilient Enterprise*. Cambridge, Massachusetts: The MIT Press.

Song J.-S., Zipkin P.H. 1996. Inventory control with information about supply conditions. *Management Science*, **42**:1409–1419.

Tomlin B. 2006. On the value of mitigation and contingency strategies for managing supply-chain disruption risks. *Management Science*, **52**:639–657.

Tomlin B. 2009a. Disruption-management strategies for short life-cycle products. *Naval Research Logistics*, **56**:318–347.

Tomlin B. 2009b. The impact of supply learning when suppliers are unreliable. *Manufacturing and Service Operations Management*, **11**:192–209.

Tomlin B., Snyder L.V. 2007. On the value of a threat advisory system for managing supply chain disruptions. Working paper.

Tomlin B., Wang Y. 2005. On the value of mix flexibility and dual sourcing in unreliable newsvendor networks. *Manufacturing and Service Operations Management*, **7**:37–57.

Wadecki A., Babich V., Wu O. 2010. Manufacturers' competition and subsidies to suppliers. In H. Gurnani, A. Mehrotra, S. Ray, editors, *Managing Supply Disruptions*. London: Springer-Verlag.

Wang Y., Gilland W.G., Tomlin B. 2010. Mitigiating supply risk: Dual sourcing or process improvement? *Manufacturing and Service Operations Management*, **12**:489–510.

CHAPTER FIVE

Beyond Risk: Ambiguity in Supply Chains

KARTHIK NATARAJAN, MELVYN SIM, AND CHUNG-PIAW TEO

5.1 Introduction to Risk and Ambiguity

The distinction between "risk" and "ambiguity" has been long made in influential works in economics and finance by Knight (1921), Keynes (1921) and Ellsberg (1961), among others. Risk refers to the situation where a decision maker can assign exact probabilities to randomness she faces. Ambiguity, on the other hand refers to the situation where randomness cannot be expressed in terms of exact probabilities. While traditionally risk modeling has dominated the research literature, there has been an increasing interest in ambiguity models. In this article, we review the notion of ambiguity arising from economics and finance and link it to the supply chain context. Our approach is descriptive with a goal to highlight the important aspects of this topic. One of the popular approaches to account for aversion to ambiguity is the maximin expected utility (MEU) theory developed by Gilboa and Schmeidler (1989). We review implications of this theory in a single period newsvendor setting. The newsvendor problem forms the

Handbook of Integrated Risk Management in Global Supply Chains, First Edition. Edited by Panos Kouvelis, Lingxiu Dong, Onur Boyabatli, and Rong Li.

foundation of many operations management models where a manager needs to decide on the order quantity before knowing the true demand. While the basic model assumes that the newsvendor knows the distribution of demand, this is often impractical. We revisit a model proposed by Scarf (1958) that accounts for imperfect demand distribution information and discuss extensions of this model. Then, we consider a supply chain inventory positioning problem that integrates the newsvendor model with a transportation model. In this case, the demand ambiguity occurs across different retail locations. Finally, we highlight possible future research directions in this area.

Expected utility (EU) theory was the earliest approach developed to account for risk in decision making under uncertainty. This theory was proposed by Bernoulli (1954) and formally axiomatized by von Neumann and Morgenstern (1944). Formally, let $\tilde{x}$ denote a random variable defined on the probability space $(\Omega, \mathcal{F}, P)$. The preference of a decision maker for the random variable $\tilde{x}$ over the random variable $\tilde{y}$ is denoted as $\tilde{x} \succeq \tilde{y}$. Throughout this section, our interpretation is that larger values are preferable to smaller values (e.g., profit). EU theory postulates that the preference relationship for a decision maker is representable with utility function $u(\cdot)$ where:

$$\tilde{x} \succeq \tilde{y} \Longleftrightarrow \mathbf{E}_P[u(\tilde{x})] \geq \mathbf{E}_P[u(\tilde{y})]$$

The expected value is calculated over the probability distribution of the outcomes P. Risk preferences of decision makers is modeled through appropriate choices of utility functions. A risk neutral, risk averse, or risk seeking decision maker is modeled through a linear, concave, or convex utility function. For strictly increasing utility functions, the preferences can be expressed using the certainty equivalent where:

$$\tilde{x} \succeq \tilde{y} \Longleftrightarrow CE_u(\tilde{x}) \geq CE_u(\tilde{y}) \;\text{ with }\; CE_u(\tilde{x}) = u^{-1}\left(\mathbf{E}_P[u(\tilde{x})]\right)$$

The certainty equivalent is interpreted as the sure amount for which the decision maker is indifferent between the sure amount and the random outcome. A related popular measure for the comparison of random outcomes for risk averse decision makers is Conditional Value-at-Risk. Conditional Value-at-Risk measures the average value of the random outcome conditional on it falling below a prespecified quantile level $\eta \in (0, 1]$. It is computed as (see Rockafellar and Uryasev (2000)):

$$CVaR_\eta(\tilde{x}) = \max_v \left(v + \frac{1}{\eta}\mathbf{E}_P\left[\min(\tilde{x} - v, 0)\right] \right)$$

For $\eta = 1$, Conditional Value-at-Risk reduces to the expected value, namely the objective function of a risk neutral decision maker. Smaller values of η corresponds to greater risk aversion. From a computational perspective, risk measures such

Acknowledgements We would like to acknowledge two former students, Joline Uichano and Song Miao, for their invaluable assistance in research related to this chapter. Joline Uichanco helped us generalize Scarf's newsvendor results while Song Miao helped us develop the supply chain inventory positioning model. We are also grateful to two anonymous reviewers for their valuable comments.

as the Conditional Value-at-Risk possess nice convexity-preserving properties, making it more amenable for large-scale optimization.

5.1.1 ORIGINS OF AMBIGUITY

A fundamental assumption under EU theory is that the decision maker knows the probability distribution of the random outcomes. In simple games of chance such as roulette, this assumption is valid for gamblers with a basic knowledge of probability. However, what happens when there is insufficient information to know the probability distribution of the random outcomes, assuming that it even exists? Subjective expected utility (SEU) theory, promoted by Savage (1954) and Anscombe and Aumann (1963) answers this question by postulating that the decision maker uses a subjective probability distribution to evaluate outcomes. In games such as horse-racing and football betting, SEU theory becomes very relevant. However Ellsberg paradox Ellsberg (1961) is a classical experiment that challenges these theories.

5.1.1.1 Ellsberg Paradox (Two-Color Experiment). There are two urns, each containing a total of 100 red and black balls. Urn I contains 50 red balls and 50 black balls. Urn II contains an unknown ratio of the two. Subjects are given four pairs of gambles to evaluate:

(1) Draw a red ball from Urn I or a black ball from Urn I
(2) Draw a red ball from Urn II or a black ball from Urn II
(3) Draw a red ball from Urn I or a red ball from Urn II
(4) Draw a black ball from Urn I or a black ball from Urn II

For each of these pairs of gambles, the subject receives \$100 if the actual draw matches the chosen option and \$0 otherwise. Experimentally it is observed that subjects are generally indifferent between the pairs of gambles in (1) and (2). For the pairs of gambles in (3) and (4), most subjects prefer drawing a red ball from Urn I over a red ball from Urn II and drawing a black ball from Urn I over a black ball from Urn II. Mathematically, let us represent the subjective probabilities of drawing a red ball from Urn II be p_r and drawing a black ball from Urn II be p_b. The set of subjective probability distributions for the balls in Urn II is:

$$\mathbb{P} = \left\{ (p_r, p_b) \,\middle|\, p_r + p_b = 1,\ p_r, p_b \geq 0 \right\}$$

For the observed preferences to be consistent with subjective utility theory, they must satisfy:

$$\frac{1}{2}u(\$100) + \frac{1}{2}u(\$0) > p_r u(\$100) + p_b u(\$0)$$

and

$$\frac{1}{2}u(\$100) + \frac{1}{2}u(\$0) > p_b u(\$100) + p_r u(\$0)$$

Assuming $u(\$100) > u(\$0) \geq 0$ (the subject strictly prefers \$100 to nothing), these relations reduce to

$$p_r < 0.5 \text{ and } p_r > 0.5$$

which are inconsistent with each other. Thus, irrespective of the specific utility function, the preferences contradict utility theory. The Ellsberg paradox helps make an important distinction between the attitude of decision makers to events with known probability (risk) and unknown probability (ambiguity). This distinction was first highlighted by Knight (1921) and Keynes (1921). More recently, these distinct notions have started influencing research in finance and economics.[1] The Ellsberg paradox provides strong evidence for ambiguity aversion—a preference for a clear bet over a vague bet. This has been substantiated in experiments in medicine and finance (see Camerer and Weber (1992)).

One popular decision-theoretic explanation to the Ellsberg paradox is MEU theory developed by Gilboa and Schmeidler (1989). Under MEU theory, a decision maker accounts for the worst possible distribution while making a decision. For the preferences in the two-color experiment to be consistent under MEU theory, it must satisfy:

$$\frac{1}{2}u(\$100) + \frac{1}{2}u(\$0) > \min_{(p_r,p_b)\in\mathbb{P}} \Big(p_r u(\$100) + p_b u(\$0)\Big)$$

and

$$\frac{1}{2}u(\$100) + \frac{1}{2}u(\$0) > \min_{(p_r,p_b)\in\mathbb{P}} \Big(p_b u(\$100) + p_r u(\$0)\Big)$$

This is equivalent to allowing for Urn II to contain 0 red balls and 100 black balls in gamble (3) and 0 black balls and 100 red balls in gamble (4). Thus, MEU theory provides a valid explanation to the Ellsberg paradox. There are two immediate predictions under this model. People approach a decision with ambiguous probabilities by considering the worst possible outcome associated with the emotional part of the brain, and yet they screen through the various options with calculated risk associated with the computational part of the brain. Recent experimental results using functional magnetic resonance imaging techniques (see Hsu et. al. (2005)) support these predictions by showing that different parts of the brain are activated when we evaluate ambiguous and risky choices. Furthermore, the research in Hsu et. al.(2005) showed that several parts of the brain (including some normally associated with the emotional side of decision making) are more

[1] As on October 6, 2009, there are around 623 articles indexed in Google Scholar between 2000 and 2009 in the Business, Administration, Finance, and Economics area with "ambiguity" in the title. There are around 533 articles indexed with the same theme between 1980 and 1999.

active under conditions of ambiguity, suggesting that the decision-making process in the brain involves integration of emotional and computational components.

5.1.2 MODELS OF AMBIGUITY

The following two issues arise in developing models of ambiguity in practical situations:

1. Characterization of the set of probability distributions $\mathbb{P}$ and
2. Characterization of the ambiguity attitude of decision-makers.

5.1.2.1 Characterization of Set of Probability Distributions. In controlled experiments such as the Ellsberg paradox and its variants (see Camerer and Weber (1992), Fox and Tversky (1995), Halevey (2007), and references therein), a small number of outcomes are provided for the subjects to evaluate. However in practical decision-making situations, one has to deal with high-dimensional joint distributions of random data with possibly an infinite number of outcomes. Examples include random returns in portfolio selection problems and random demands in supply chain problems. Characterizing the set of probability distributions $\mathbb{P}$ is then a challenging task in itself. Simple and useful characterizations that can account for possible dependencies among the random terms becomes particularly relevant. Historical data and/or expert estimates serves as tools to guide in the estimation of $\mathbb{P}$. In this context, *mean-variance* based characterizations of ambiguity with *descriptive statistics* to supplement the information can be a simple, yet rich tool to describe $\mathbb{P}$. Such characterizations of ambiguity are convenient in terms of estimation and optimization. In the portfolio selection problem, Garlappi et al. (2007), Popescu (2007), Natarajan et al. (2010a), and Delage and Ye (2010) develop optimal solutions under the worst-case approach with (partial) information on the mean and covariance matrix of the returns. No assumptions on the exact form of the distribution are made in the results therein, thus accounting for non-normality in the return data. However, by using only the mean and covariance information, any known skewness information on the data is neglected. Natarajan et al. (2010a, 2010b) incorporate information on the skewness of distributions through the use of partitioned statistics such as semi-variance. In the supply chain context, Scarf (1958) as early as 1958, characterized the optimal order quantity for a risk neutral newsvendor under MEU theory. His result was based on the assumption that the mean and variance of the demand is known but the exact form of the distribution is unknown. We discuss his model and its generalizations in the later sections, arguing that it is a useful approach to characterize ambiguity in supply chain problems.

5.1.2.2 Characterization of Ambiguity Attitude. From a behavioral perspective, other decision criteria have been proposed to address varying attitudes to ambiguity. The preference relationship under MEU theory can be represented as:

$$\tilde{x} \succeq \tilde{y} \Longleftrightarrow \min_{P\in\mathbb{P}} \mathbf{E}_P[u(\tilde{x})] \geq \min_{P\in\mathbb{P}} \mathbf{E}_P[u(\tilde{y})]$$

While this helps explain the results from the Ellsberg experiment, there are instances when a decision maker might prefer an ambiguous situation to a less ambiguous situation. Ellsberg himself noted that in his experiment, decision makers could prefer ambiguous alternatives when the probability of gains were low and the probability of losses were high (see Einhorn and Hogarth (1986) for a detailed discussion on this topic). In another set of interesting experiments, Heath and Tversky (1991) argued that the competence and expertise of the subjects played a role in their attitude towards ambiguity. Subjects who were knowledgable about sports but not politics preferred to bet on sports events rather than chance events that these people had judged to be equally probable. However, when the same subjects were asked to bet on political events, they preferred the chance event over the political event that they had perceived to be equally probable. People who were knowledgable about politics but not sports exhibited the reverse pattern. Clearly, this is not consistent with ambiguity aversion since the chance events as constructed were inherently less ambiguous than the actual events themselves. Such experimental results have led to extensions of MEU theory. Two of the popular extensions are:

1. α-Maximin Expected Utility (α-MEU). Ghirardato et al. (2004) propose that the preference is characterized by a single parameter $\alpha \in [0, 1]$. The extreme value $\alpha = 1$ models ambiguity aversion while $\alpha = 0$ models ambiguity seeking. Values of $\alpha \in (0, 1)$ trade-off between pessimism and optimism. The preference relationship $\tilde{x} \succeq \tilde{y}$ under this theory is given as:

$$\alpha \min_{P\in\mathbb{P}} \mathbf{E}_P[u(\tilde{x})] + (1-\alpha) \max_{P\in\mathbb{P}} \mathbf{E}_P[u(\tilde{x})] \geq \alpha \min_{P\in\mathbb{P}} \mathbf{E}_P[u(\tilde{y})] + (1-\alpha) \max_{P\in\mathbb{P}} \mathbf{E}_P[u(\tilde{y})].$$

2. Smooth Recursive Expected Utility. Klibanoff et al. (2005) propose a smoothed model that accounts for not just the worst and the best case distributions, but for all possible distributions. Their preference is based on a probability distribution $Q(P)$ that is a subjective probability measure on the relevance of each probability distribution $P \in \mathbb{P}$. A function $\phi(\cdot)$ captures the ambiguity preference while $u(\cdot)$ captures the risk preference, thus differentiating the two concepts. The preference relationship $\tilde{x} \succeq \tilde{y}$ under this theory is given as:

$$\tilde{x} \succeq \tilde{y} \Longleftrightarrow \mathbf{E}_{Q(P)}[\phi\left(\mathbf{E}_P[u(\tilde{x})]\right)] \geq \mathbf{E}_{Q(P)}[\phi\left(\mathbf{E}_P[u(\tilde{y})]\right)]$$

In the next two sections, we focus on the ambiguity averse MEU theory. Analyzing the relevance and implications of the other models in supply chain problems is an important question that is left for future research.

5.2 Ambiguity in a Single Period Newsvendor Setting

One of the fundamental problems in operations management is the newsvendor problem. In this section, we review a model pioneered by Scarf (1958) that addresses the issue of ambiguity in the newsvendor problem. We highlight an extension of the result from the risk neutral to the risk averse newsvendor and provide a numerical example to validate the usefulness of the approach.

Consider a newsvendor who sells a seasonal product, say newspapers. She places an order with the supplier, before knowing the actual demand. The order quantity is used to satisfy as much of the demand as possible. Any unmet demand is lost while any excess amount is salvaged at a value of zero. Let c be the unit ordering cost and $p > c$ be the unit selling price. For an order quantity q, the profit for the newsvendor under the random demand $\tilde{d}$ is:

$$\pi(q, \tilde{d}) = p \min(q, \tilde{d}) - cq$$

Suppose the newsvendor wants to find an order quantity to maximize the expected profit conditional on it falling below a quantile level $\eta \in (0, 1]$. She would use a Conditional Value-at-Risk criterion and solve the following problem:

$$\max_{q \geq 0} CVaR_\eta \left(\pi(q, \tilde{d})\right)$$

Assuming that demand is continuously distributed with a cumulative distribution function $F(x) = P(\tilde{d} \leq x)$, the optimal order quantity is:

$$q^* = F^{-1}(\eta\beta)$$

where $\beta = 1 - c/p$.

The proof of this result can be found in Gotoh and Takano (2007) and Chen et al. (2009). The corresponding optimal objective value is:

$$CVaR_\eta \left(\pi(q^*, \tilde{d})\right) = \frac{1}{\eta} \int_0^{F^{-1}(\eta\beta)} pxdF(x)$$

For $\eta = 1$, this reduces to a risk neutral newsvendor who is only interested in maximizing her expected profit. This leads to the well-known critical fractile solution of $F^{-1}(\beta)$.

5.2.1 SINGLE PERIOD NEWSVENDOR: MAXIMIN EXPECTED UTILITY

One of the key inputs to the newsvendor model is the demand distribution. Demand models for the newsvendor are generated using historical data and/or subjective forecasting methods. In using historical data, a newsvendor is typically concerned about the possibility that the future demand many not be identical to the past demand data. Furthermore, when introducing new products to new markets, it is difficult to have access to historical data. Even with sophisticated forecasting techniques, a newsvendor at best might believe in a set of demand models, rather than a single model.

Consider a newsvendor who assumes that the demand distribution lies in a set $\mathbb{P}$, but does not believe in any particular distribution. A newsvendor under the MEU theory would take a conservative approach by maximizing the worst-case Conditional Value-at-Risk over the set of distributions. This is equivalent to solving the following:

$$\max_{q\geq 0}\min_{P\in\mathbb{P}} CVaR_\eta\left(\pi(q,\tilde{d})\right)$$

A choice for $\mathbb{P}$ is the set of non-negative demand distributions with a known mean value μ and standard deviation σ. This characterization of ambiguity was first proposed by Scarf (1958). In the risk neutral case, he provided an explicit expression for the optimal order quantity. Formally, the set of distributions is defined as:

$$\mathbb{P} = \left\{P \;\middle|\; P[\tilde{d}\geq 0] = 1,\ \mathbf{E}_P[\tilde{d}] = \mu,\ \mathbf{E}_P[\tilde{d}^2] = \mu^2+\sigma^2\right\}$$

No assumptions on the specific shape of the demand distribution is made. Features such as symmetry, asymmetry, unimodality can be incorporated into this model through the use of semidefinite programming (see Perakis and Roels (2008), Natarajan et al. (2010b), Bertsimas and Popescu (2002)). The next result provides the generalization of Scarf's result to the risk averse newsvendor.

Proposition 5.1 *(Natarajan et al. (2010b)) The optimal order quantity for a risk averse newsvendor under the MEU approach with a η-Conditional Value-at-Risk criterion given the set of demand distributions with mean μ and standard deviation $\sigma > 0$ is:*

$$q^* = \begin{cases} 0 & \text{if } \eta\beta < \frac{\sigma^2}{\mu^2+\sigma^2}, \\ \mu + \frac{\sigma}{2}\left(\frac{2\eta\beta-1}{\sqrt{\eta\beta(1-\eta\beta)}}\right) & \text{otherwise.} \end{cases}$$

Under this order quantity, the worst-case Conditional Value-at-Risk is:

$$\min_{P\in\mathbb{P}} CVaR_\eta\left(\pi(q^*,\tilde{d})\right) = \begin{cases} 0 & \text{if } \eta\beta < \frac{\sigma^2}{\mu^2+\sigma^2}, \\ (p-c)\left(\mu - \sigma\sqrt{\frac{1-\eta\beta}{\eta\beta}}\right) & \text{otherwise.} \end{cases}$$

For $\eta = 1$, the result reduces to the result of Scarf (1958). For a fixed β, as the aversion to risk increases (or η decreases), the optimal order quantity clearly decreases. This is consistent with the result for risk averse newsvendors obtained in Eeckhoudt et. al. (1995). A natural extension of this result when the parameters of the demand distribution are themselves unknown is provided next. For the risk neutral newsvendor, the result reduces to the order quantity in Kouvelis and Yu (1997).

Corollary 5.1 *The optimal order quantity for a risk averse newsvendor under the MEU approach with a η-Conditional Value-at-Risk criterion over the set of demand distributions with mean $\mu \in [\underline{\mu}, \overline{\mu}]$ and standard deviation $\sigma \in [\underline{\sigma}, \overline{\sigma}]$ is:*

$$q^* = \begin{cases} 0 & \text{if } \eta\beta < \frac{\overline{\sigma}^2}{\underline{\mu}^2 + \overline{\sigma}^2}, \\ \underline{\mu} + \frac{\overline{\sigma}}{2}\left(\frac{2\eta\beta - 1}{\sqrt{\eta\beta(1-\eta\beta)}}\right) & \text{otherwise.} \end{cases}$$

The next numerical example, compares the performance of the MEU approach with a conventional sampling based approach for the newsvendor problem. Consider a risk neutral newsvendor who minimizes the worst-case sum of the expected overage and underage costs where h and b are the unit overage and underage costs respectively. This problem is the cost minimization version of the profit maximization problem considered thus far. In this example, we used demand samples that were artificially mapped from a publicly available financial data set. The data was based on daily returns of the Fama & French[2] 49 industry portfolio. The portfolio consisted of NYSE, AMEX and NASDAQ stocks classified by industry. These included industries such as finance, health, textiles, food, and machinery. We used a total of 1259 daily returns spanning a total of five years, from January 2, 2004 to December 31, 2008. The daily demands of each retailer were mapped to the daily returns of the Fama & French 49 industry portfolios in an affine manner, so that the demands were non-negative and preserved the volatility of the underlying returns. The demand of retailer i at time t was set as:

$$d_{it} = 100 \times r_{it} + \phi$$

where r_{it} is the t-th day return of the i-th industry and $\phi = 2000$.

The industries were assumed to correspond to demands faced by different retailers. The parameter ϕ was chosen so that the demands were non-negative. Figure 5.1 presents the average daily demands and standard deviations across the retailers. The figure clearly shows that the standard deviations increased significantly in the year 2008 with a sharp drop in average demands. This corresponds to the period of the global financial crisis.

The experiments were conducted as follows. The time period was started at $t = 250$, which is approximately the beginning of the year 2005. Using the demand information for 250 days of the preceding year (i.e., from $t = 1$ through $t = 250$) we computed the order quantity via two approaches. The first is the sampling approach, which minimized the average cost using the historical demand data:

$$\min_{q \geq 0} \sum_{t=1}^{250} \frac{1}{250} \left[h(q - d_{it})^+ + b(d_{it} - q)^+ \right]$$

[2] Data is from http://mba.tuck.dartmouth.edu/pages/faculty/ken.french/data_library.html

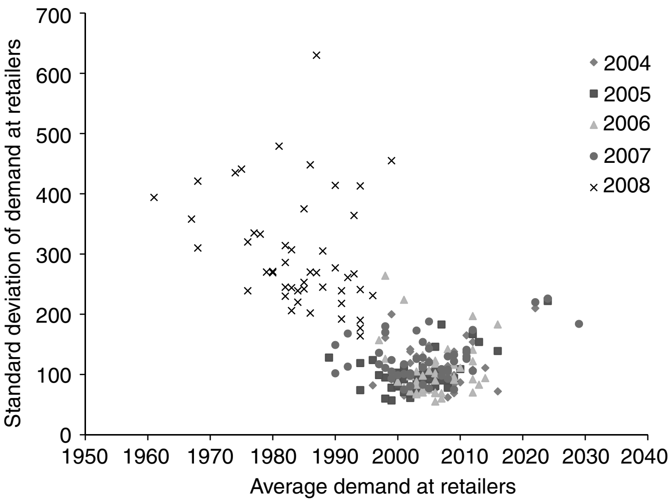


FIGURE 5.1 **Sample mean and standard deviation across retailers across years.**

The second is the MEU approach based on the sample mean and variances computed from the same set of data:

$$\min_{q\geq 0}\max_{P\in\mathbb{P}}\mathbf{E}_P\left[h(q-\tilde{d}_i)^+ + b(\tilde{d}_i-q)^+\right]$$

After determining the ordering quantities, we computed the actual average daily costs using the realized demand information from $t+1$ to $t+21$, which is approximately the length of a month. For simplicity, we assumed that excess inventory is not carried over to the next day. Subsequently, we advanced to time period $t+21$ and repeated the process again. After 12 cycles, which is approximately the end of the year 2005, the average daily costs for the two approaches were compared. In an identical manner, results for the years 2006 through 2008 were generated.

In the numerical experiments, we observed that performance of both approaches were similar when the disparity between the overage and underage costs was small ($1/5 \leq b/h \leq 5$). When the disparity of the holding and shortage costs was greater, the results were less similar. Figures 5.2 and 5.3 display the results for the case of $b = 1$, $h = 20$ and $b = 20$, $h = 1$. During the period from 2004 through 2006 in which the demand distribution was fairly stable, the performance of the sample approach was slightly better than the MEU approach. In the year 2007, the MEU approach fared slightly better than the sample approach. In the year 2008, when the fluctuations of demands deviated significantly from the past data, MEU significantly outperformed the sample approach. As this example illustrates, the performance of the MEU approach seems to be good when there is greater uncertainty in the future demand distribution. In this case, this corresponds to the global financial crisis.

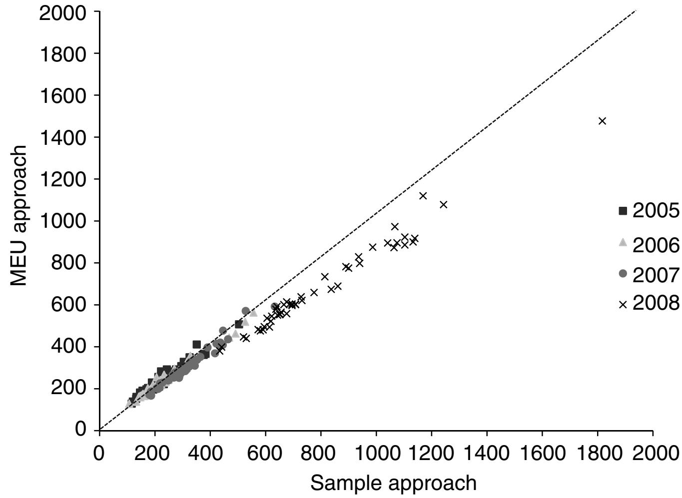


FIGURE 5.2 Comparison of expected costs for the MEU and sample approach: b = 1, h = 20.

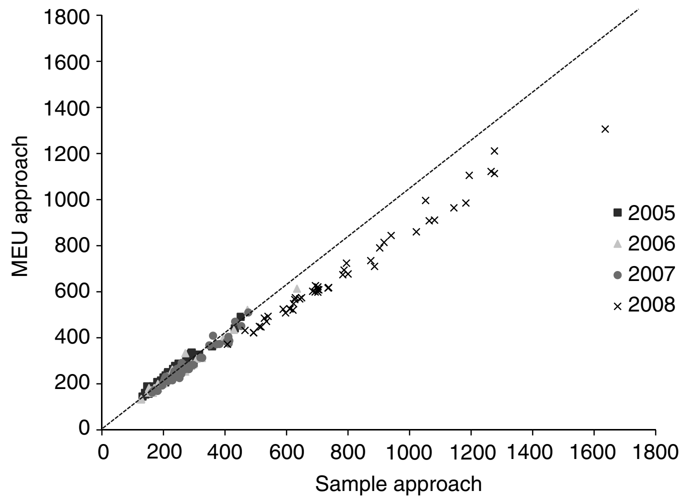


FIGURE 5.3 Comparison of expected costs for the MEU and sample approach: b = 20, h = 1.

5.3 Ambiguity in a Supply Chain Inventory Positioning Setting

In this section, we study the effects of ambiguity in a supply chain network involving multiple supply and demand locations. Using a stochastic programming framework, we propose a tractable model for a decision maker who must decide on the transportation quantities before knowing the actual demand distribution. The model generalizes Scarf's single newsvendor problem to the multiple location setting.

In the annals of the history of business there are abundant examples of big companies devastated by unexpected events. Enron was widely praised as a model of sound management, until its dramatic collapse due to accounting irregularities. This soon led to the creation of the Basel II accord in 2004 to develop an international standard for banking regulators to use in determining the the amount of capital needed to guard against different types of financial and operational risks to prevent insolvency. This development has also led to a surge in the research activities on the topic of supply chain risk management. Based on a sample of 519 glitches announcements made during 1989–2000, Hendricks and Singhal (2003) noted that shareholder return typically falls by 7% to 8% on the day a disruption is announced, whereas operating income falls by 42%. The effect is even more pronounced during a natural disaster, such as the 1995 earthquake in Kobe, which destroyed the Port of Kobe and a significant part of the world's electronic industry, which was sited nearby. While such external risks are normally outside manager's control, characterizing or quantifying the risks associated with such rare events is also nearly impossible. Instead, companies have opted to cope with external risks with process solutions such as business continuity planning, crisis monitoring, and/or recovery management teams.

On the other end of the spectrum, internal risks in the form of late deliveries, wrong forecast, human errors, breakdown of equipments, and so on are generally less dramatic, but are more frequent and widespread in their impact on supply chain performance. The cascading effect of disruption down the supply chain can often exacerbate its impact on information distortion. To deal with these risks, companies have to build capability to plan and operate in uncertain environment—making the task of dealing with ambiguity in the supply chain extremely challenging. Several quantitative approaches have been proposed to try and find solutions that works reasonably well under a variety of scenarios in supply chain problems. Bertsimas and Thiele (2006) and the references therein propose an approach using robust optimization while Shapiro and Kleygweyt (2002) propose a minimax approach using the method of sampling average approximation. We next evaluate a methodology to incorporate the effects of ambiguity in a simple supply chain network using MEU theory with mean and variance demand information.

Consider a set of suppliers denoted by $\mathcal{S}$ and a set of retailers denoted by $\mathcal{R}$. Each supplier $i \in \mathcal{S}$ provides fixed s_i units of the product and each retailer $j \in \mathcal{R}$ faces an uncertain demand $\tilde{d}_j$ for the product. For each unit of unsatisfied demand at retailer j there is a penalty cost b_j and for each unsold unit there is an inventory holding cost h_j. The unit transportation cost from supplier i to retailer j is c_{ij}. The stochastic supply chain inventory positioning problem is to find the amount of product q_{ij} to be transported from supplier i to retailer j before the actual value of the demand is known so as to minimize costs. We restrict ourselves to a nondynamic setting where the holding costs are used to reflect interactions with future periods. This problem is common in supply chain planning and integrates the newsvendor model with the transportation model. For example, if $\mathcal{R}$ denotes a set of retail outlets in the newspaper distribution business and $\mathcal{S}$ denotes a set of printing facilities, the goal is find the optimal plan to distribute all the newspapers to the outlets, minimizing the total transportation and newsvendor costs. This problem was first studied by Williams (1963) who formulated it as the following two-stage stochastic program:

$$\begin{aligned}
\min \ & \sum_{i\in\mathcal{S}}\sum_{j\in\mathcal{R}} c_{ij}q_{ij} + \mathbf{E}_P\left[\mathcal{Q}(\boldsymbol{q},\tilde{\boldsymbol{d}})\right] & \\
\text{s.t.} \ & \sum_{j\in\mathcal{R}} q_{ij} = s_i & \forall i \in \mathcal{S} \\
& q_{ij} \geq 0 & \forall i \in \mathcal{S}\,\forall j \in \mathcal{R}
\end{aligned}$$

The second-stage cost in this problem is computed as:

$$\begin{aligned}
\mathcal{Q}(\boldsymbol{q},\tilde{\boldsymbol{d}}) = \min \ & \sum_{j\in\mathcal{R}} b_j q_j^+ + \sum_{j\in\mathcal{R}} h_j q_j^- & \\
\text{s.t.} \ & q_j^+ - q_j^- = \tilde{d}_j - \sum_{i\in\mathcal{S}} q_{ij} & \forall j \in \mathcal{R} \\
& q_j^+, q_j^- \geq 0 & \forall j \in \mathcal{R}
\end{aligned}$$

The variables q_j^+ and q_j^- in the second-stage problem denote the quantity of unsatisfied demand and unsold amount at retailer j. The overall objective of the decision maker is to minimize the sum of the first-stage transportation costs $\sum_{i,j} c_{ij}q_{ij}$ and the expected second-stage penalty and holding costs $\mathbf{E}_P[\mathcal{Q}(\boldsymbol{q},\tilde{\boldsymbol{d}})]$. The special structure of the second-stage problem in this case implies that the optimal decisions depends only on the marginal distribution of the demand at each of the retailers. If the marginal distribution of the demand $\tilde{d}_j$ at the retailer j is P_j, the stochastic program can be reformulated as:

$$\begin{aligned}
\min \ & \sum_{i\in\mathcal{S}}\sum_{j\in\mathcal{R}} c_{ij}q_{ij} + \sum_{j\in\mathcal{R}} \mathbf{E}_{P_j}\left[\max\left(b_j\left(\tilde{d}_j - \sum_{i\in\mathcal{S}} q_{ij}\right), h_j\left(\sum_{i\in\mathcal{S}} q_{ij} - \tilde{d}_j\right)\right)\right] & \\
\text{s.t.} \ & \sum_{j\in\mathcal{R}} q_{ij} = s_i & \forall i \in \mathcal{S} \\
& q_{ij} \geq 0 & \forall i \in \mathcal{S}\,\forall j \in \mathcal{R}
\end{aligned}$$

5.3.1 SUPPLY CHAIN INVENTORY POSITIONING: MAXIMIN EXPECTED UTILITY

Instead of assuming a single demand distribution at each of the retailers, let $\mathbb{P}_j$ represent the set of possible distributions at each retailer $j \in \mathcal{R}$. Suppose the demand at retailer j is assumed to lie in the range $[\underline{d}_j, \overline{d}_j]$ with mean μ_j and variance σ_j^2. The sets of distributions are characterized as:

$$\mathbb{P}_j = \left\{ P_j \;\middle|\; P_j(\underline{d}_j \le \tilde{d}_j \le \overline{d}_j) = 1,\ \mathbf{E}_{P_j}[\tilde{d}_j] = \mu_j,\ \mathbf{E}_{P_j}[\tilde{d}_j^{\,2}] = \mu_j^2 + \sigma_j^2 \right\} \forall j \in \mathcal{R}$$

A decision maker under the MEU approach would choose the transportation quantities to minimize the worst-case expected costs across all possible distributions of the demands at the retailers. The mathematical formulation of this problem is:

$$\begin{aligned}
\min \; & \sum_{i\in\mathcal{S}} \sum_{j\in\mathcal{R}} c_{ij} q_{ij} + \sum_{j\in\mathcal{R}} \max_{P_j \in \mathbb{P}_j} \mathbf{E}_{P_j} \left[\max \left(b_j \left(\tilde{d}_j - \sum_{i\in\mathcal{S}} q_{ij} \right), h_j \left(\sum_{i\in\mathcal{S}} q_{ij} - \tilde{d}_j \right) \right) \right] \\
\text{s.t.} \; & \sum_{j\in\mathcal{R}} q_{ij} = s_i \qquad \forall i \in \mathcal{S} \\
& q_{ij} \ge 0 \qquad \forall i \in \mathcal{S}\, \forall j \in \mathcal{R}
\end{aligned}$$

In this case, the problem can be reformulated as a second-order conic program. For details on converting the formulation to a second-order conic program, the reader is referred to Nesterov (2000), Bertsimas and Popescu (2002) and Natarajan et al. (2009). Second-order conic programs can be efficiently solved using interior point methods.

Our first example provides a simple illustration on the importance of accounting for demand variability in this problem. Consider the one supplier, two retailer setting displayed in Figure 5.4. The supplier has a fixed supply rate of 2. Retailer 1 faces one-unit demand and retailer 2 faces stochastic demand with rate uniformly distributed in the range [0.5, 1.5]. The transportation costs from the supplier to the retailers are assumed to be 0. All other costs are indicated in the figure. Obviously, if the random demand is replaced by the expected value, we should ship one unit to each retailer with an expected total cost of 1.2625. This is referred to as the expected value solution. The expected value solution can be improved by increasing the shipment to retailer 2. The optimal solution to the stochastic program is in fact to transport 1.4802 units to retailer 2 and the rest of the supplies to retailer 1. In this case, the expected total cost is reduced to

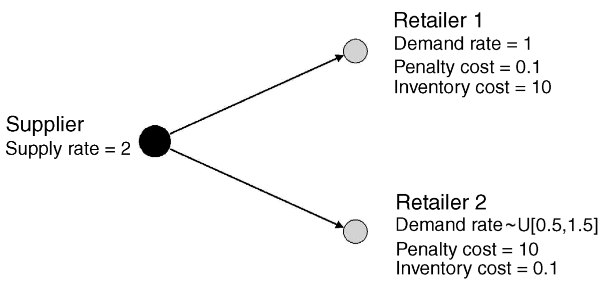


FIGURE 5.4 **A one supplier, two retailer example.**

0.0980. The ability to account for the random demand in the model improves the objective value substantially. The optimal solution under the MEU approach with mean, variance, and support information is to transport 0.5 units to retailer 1 and 1.5 units to retailer 2. The expected cost under this solution is 0.1, which is much better than the cost from the expected value solution and close to the stochastic programming solution.

Our next example is a large-scale illustration to demonstrate the effects of incorporating ambiguity into the model. The experiments were run on a personal computer with an Intel Pentium 4 M, 1.72 GHz CPU, 256 MB of RAM and Microsoft Windows XP professional operating system. The codes were run in MATLAB 6.5 using SeDuMi 1.05R5 as the solver for the optimization problems. The experiments were run on 20 randomly generated instances with both $|\mathcal{S}|$ and $|\mathcal{R}|$ set to 10, 20, 50, 100, and 200 respectively. All cost components (the shortage costs, inventory holding costs, and transportation costs) were generated uniformly and independently in the range [0, 1]. The supply quantities were generated uniformly in the range [0.5, 1.5]. The lower bound $\underline{d}_j$ and the upper bound $\overline{d}_j$ of each demand was generated uniformly in the range [0, 1] and [1, 2] respectively. The mean $\mu_j = (\underline{d}_j + \overline{d}_j)/2$ and the variance $\sigma_j^2 = (\overline{d}_j - \underline{d}_j)^2/12$ were set based on a uniform distribution. Three approaches were tested on this problem:

1. Deterministic or expected value approach. The expected value of the demand was used as a single scenario in finding an optimal solution. This reduces to solving the following small-sized linear program:

$$
\begin{aligned}
\min \ & \sum_{i\in\mathcal{S}}\sum_{j\in\mathcal{R}} c_{ij}q_{ij} + \sum_{j\in\mathcal{R}}\left(b_j q_j^+ + h_j q_j^-\right) & \\
\text{s.t. } & \sum_{j\in\mathcal{R}} q_{ij} = s_i & \forall i \in \mathcal{S} \\
& q_j^+ - q_j^- = \mu_j - \sum_{i\in\mathcal{S}} q_{ij} & \forall j \in \mathcal{R} \\
& q_{ij} \geq 0 & \forall i \in \mathcal{S}\, \forall j \in \mathcal{R} \\
& q_j^+, q_j^- \geq 0 & \forall j \in \mathcal{R}.
\end{aligned}
$$

2. Stochastic programming approach. A sampling based technique was used to solve the stochastic program with 20 and 50 samples generated independently from the uniform distribution for each demand. While special-purpose algorithms can be developed for uniform distributions (see Williams (1963)), we use a general purpose linear programming approach to allow for the fact that demands need not always fit simple distributions. The corresponding large-scale linear program with T samples for the demand at each retailer is:

$$\begin{aligned}
\min \sum_{i\in\mathcal{S}}\sum_{j\in\mathcal{R}} c_{ij}q_{ij} + \frac{1}{T}\sum_{j\in\mathcal{R}}\sum_{t=1}^{T}\left(b_j q_{jt}^{+} + h_j q_{jt}^{-}\right) & \\
\text{s.t.} \sum_{j\in\mathcal{R}} q_{ij} = s_i & \qquad \forall i \in \mathcal{S} \\
q_{jt}^{+} - q_{jt}^{-} = d_{jt} - \sum_{i\in\mathcal{S}} q_{ij} & \qquad \forall j \in \mathcal{R}\ \forall t = 1,\ldots,T \\
q_{ij} \geq 0 & \qquad \forall i \in \mathcal{S}\ \forall j \in \mathcal{R} \\
q_{jt}^{+}, q_{jt}^{-} \geq 0 & \qquad \forall j \in \mathcal{R}\ \forall t = 1,\ldots,T
\end{aligned}$$

3. MEU approach. Given the mean, variance, and support information on the demand, the MEU approach reduces to solving a small-sized second-order conic program as discussed earlier.

The computational times for the three different approaches are shown in Figure 5.5. An upper bound of 180 seconds was set on the computational times for each of the approaches. The horizontal axis represents the number of suppliers (retailers), and the vertical axis corresponds to the average computational time over the 20 instances in seconds. The deterministic model has the shortest computational time. The computational time for the MEU approach is less than twice that of the deterministic model. The sample based approaches require greater computational effort with the 50 samples and 200 suppliers case exceeding the time limit.

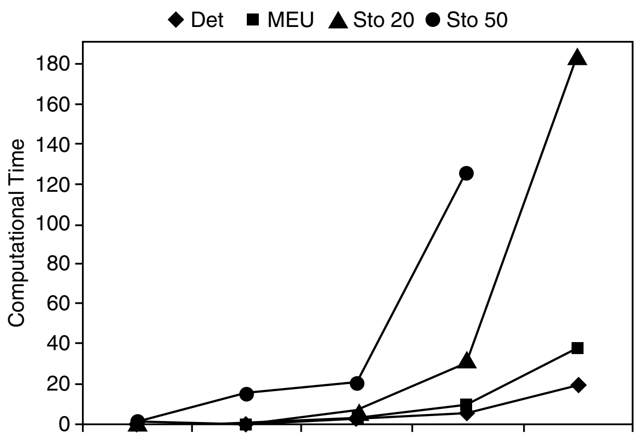


FIGURE 5.5 **Computational times for a large-scale supply chain inventory positioning problem.**

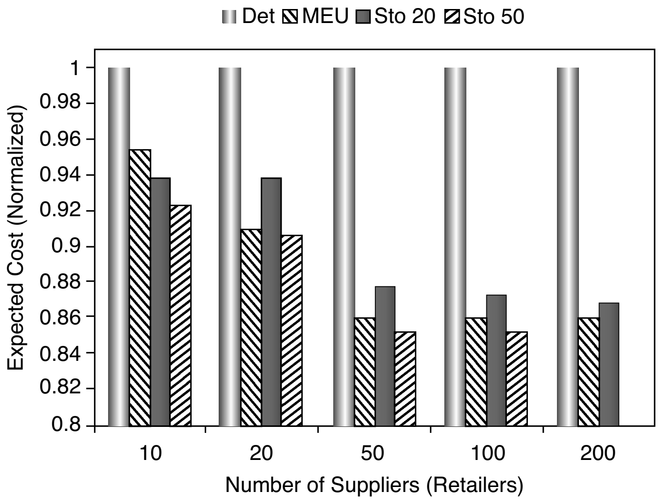


FIGURE 5.6 **Optimal expected cost under uniform distribution.**

To compare the approaches, we calculated the corresponding costs for different samples[3] if the solutions to these models were implemented. Figure 5.6 shows the average performance over 20 instances for varying number of suppliers and retailers. We divide the cost for each model by the cost computed using the deterministic model. Compared with the deterministic approach, the MEU approach has much better performance, especially for the larges-scale problems. For the 200 suppliers and retailers problems, the cost is merely 85.6% of the deterministic approach. At the same time, the approach outperforms the stochastic model with 20 samples except for the 10 suppliers (retailers) problem. Even in that case, the improvement of the stochastic model is only 1.5%. The stochastic model with 50 samples performs better than the MEU approach. However, the average improvement is 1.3%. In particular, for the 20 suppliers and retailers problem, it is only 0.3% better than the MEU approach. This indicates that to get reliable estimates from the sampling approach, one must ensure that a sufficient number of samples are available.

We next analyze the impact of mis-specifying the demand distribution. Note that demand distribution estimation error has no impact on the results under MEU as long as the range, mean and variance parameters are unchanged. However the performance of policies obtained using the classical stochastic programming framework will be affected by estimation error. What is the impact of demand estimation error on the performance? We investigate this issue next.

Suppose that stochastic programming approach is solved assuming samples from a uniform distribution, but in fact the real demand is a (truncated) normal distribution. The MEU approach was solved using the same mean and variance

[3] Over 20 samples were generated using the prescribed uniform distribution.

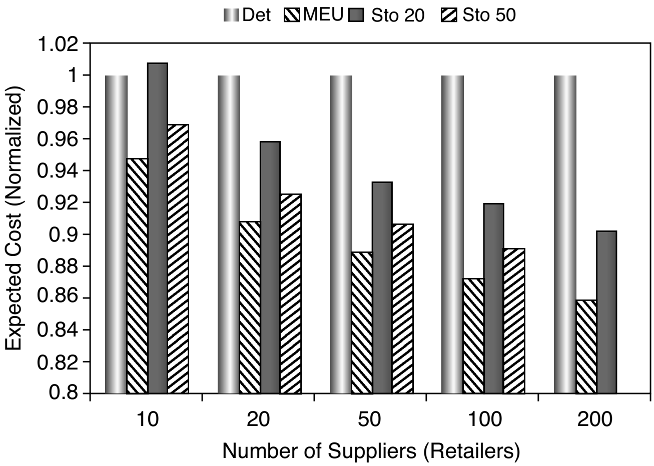


FIGURE 5.7 Optimal expected cost under mis-specified distribution.

but with the support of $\tilde{d}_j$ set to be $[\mu_j - 3\sigma_j, \mu_j + 3\sigma_j]$. Demands were sampled from a (truncated) normal distribution to assess the performance of the different models. Figure 5.7 displays the comparison between different approaches. When a normal distribution is used to to evaluate the model, the improvement from the MEU approach over the deterministic model, increases with the scale of the problems from 5.1% to 14.2%. The stochastic programming models also suffer from the estimation errors of the distribution assumption. On average, the MEU approach is 4.9% better than the 20 sample model, and 1.8% better than the 50 sample model.

5.4 Conclusions

In this chapter, we have reviewed the concept of ambiguity that has its origins in the economics and finance literature. This concept has been less explored in supply chains. Yet it is potentially useful due to the inherent ambiguity in demand information in supply chain problems. Using two simple illustrations from the newsvendor and supply chain inventory positioning context, we have highlighted techniques to account for demand ambiguity. We end this chapter by outlining possible future directions for supply chain researchers in this area:

1. *Exploring models that capture different attitudes to ambiguity:*
   The newsvendor order quantity proposed by Scarf (1958) is based on the MEU worst-case approach. In Section 5.1, we discussed two extensions of this theory that account for possibly different attitudes to ambiguity. To the best of our knowledge, the applications of these models to the newsvendor problem have been unexplored. Chateauneuf et al. (2007) showed that the α-MEU

theory, which accounts for both optimism and pessimism, can help address some economic paradoxes such as co-existence of gambling and insurance purchase, the equity premium puzzle and the small stock puzzle. In a related vein, the relevance of these theories in supply chains and their implications form an interesting future research direction.

2. *Understanding the behavior of supply chain managers under ambiguity:*
In a well-known controlled experiment, Schweitzer and Cachon (2000) showed that newsvendors often tend to order too many of low-profit products and too few of high-profit products. Furthermore, this behavior was not consistent with models such as risk aversion or risk seeking. Among the plausible explanations they provided were the mean anchoring heuristic and preferences to reduce the ex-post inventory error. Several other researches (Bolton and Katok 2008, Corbett and Fransoo 2007, Lurie and Swaminathan 2009) have confirmed the evidence that the newsvendor decisions systematically deviates from the standard prescription of the optimal order quantity. Benzion et al. (2010) performed the same set of experiments under the additional condition that half the newsvendors did not even know the demand distribution. They found that behavioral biases still persisted for both sets of newsvendors. Surprisingly, they found that the subjects who knew the demand distribution ordered on average the same as the subjects who did not know the distribution. Thus the absence of the knowledge of the exact demand distribution did not change the bias significantly. Also, the ordering policies of subjects who did not know the distribution were more affected by feedback. We believe that there is a greater need for such studies that help understand the true behavior of decision makers under ambiguity.

3. *Analyzing the computational complexity of ambiguity models:*
One of the computational challenges in using the MEU theory is the calculation of the worst possible expected objective value over all plausible joint distributions of the data. This in turn makes the computation of the optimal decision variables highly challenging. For example, by incorporating additional covariance information into the ambiguity model developed in Section 5.3, the problem becomes NP-hard (see Bertsimas et al. (2010)). From a practical perspective, this implies that efficiently solving MEU theory based models in multivariate settings is unlikely. This naturally leads to an interest in developing approximate solutions that are provably close to optimal. There is a need to understand the computational complexity of such models particularly in the multivariate setting.

4. *Studying the performance of supply chains under ambiguity:*
The topic of coordination in supply chains has been an active area of research over the past two decades (see Cachon (2003) for a detailed review of models). The research agenda in this area is to develop contracts that can ensure that all the entities in the supply chain work towards a common goal such as total supply chain profit maximization. For example, the design and analysis of the buyback contract that coordinates the performance of a manufacturer and a retailer in a two-echelon supply chain is well understood. The standard model

in this area assumes that the two entities share the same market information and demand distribution. However, even if the market information is the same there is no reason to believe that both entities would have the same assessment of the demand distribution of a new product. Also, a manufacturer is often not privy to the retailer's information (see Desiraju and Moorthy (1997)). In this context, Lau and Lau (2001) studied a two-echelon model where the retailer has superior market information. Therein, the manufacturer must decide on the wholesale price and the retailer must decide on the order quantity in the face of random demand. Their results showed that the retailers improved market knowledge helps the manufacturer and the supply chain but not necessarily the retailer himself, as long as the manufacturer knows that the retailer has superior demand information. Whether similar conclusions hold for MEU theory and more generally how different supply chain contracts perform under ambiguity is another interesting area for further research.

REFERENCES

Anscombe F.J., Aumann R.J. 1963. A definition of subjective probability. *Annals of Mathematical Statistics,* 34: 149–205.

Benzion U., Cohen Y., Shavit T. 2010. The newsvendor problem with unknown demand distribution. *Journal of the Operational Research Society,* 6(6): 1022–1031.

Bernoulli D. 1954 (originally published in 1738; translated by Dr. Lousie Sommer). Exposition of a new theory on the measurement of risk. *Econometrica,* 22(1): 22–36.

Bertsimas D., Doan X.V., Natarajan K., Teo C.P. 2010. Models for minimax stochastic linear optimization problems with risk aversion. *Mathematics of Operations Research,* 35(3): 580–602.

Bertsimas D., Popescu I. 2002. On the relation between option and stock prices: An optimization approach. *Operations Research,* 50(2): 358–374.

Bertsimas D., Thiele A. 2006. A robust optimization approach to inventory theory. *Operations Research,* 54(1): 150–168.

Bolton, G.E. Katok E. 2008. Learning by doing in the newsvendor problem: A laboratory investigation of the role of experience and feedback. *Manufacturing and Service Operations Management,* 10(3): 519–538.

Cachon G. Supply chain coordination with contracts. (2003. In: *Handbooks in Operations Research and Management Science: Supply Chain Management.* Graves S.C. De Kok A.G., editors. North Holland, Amsterdam, The Netherlands.

Camerer C., Weber M. 1992. Recent developments in modeling preferences: Uncertainty and ambiguity. *Journal of Risk and Uncertainty,* 5: 325–370.

Chateauneuf A., Eichberger J., Grant S. 2007. Choice under uncertainty with the best and worst in mind: Neo-additive capacities. *Journal of Economic Theory,* 137(4): 538–567.

Chen Y.F., Xu M., Zhang Z.G. 2009. Technical note—A risk averse newsvendor model under the CVaR criterion. *Operations Research,* 57(4): 1040–1044.

Corbett C.J., Fransoo J.C. 2007. Entrepeneurs and newsvendors: Do small businesses follow the newsvendor logic when making inventory decisions. UCLA: Working Paper.

Delage E., Ye Y. 2010. Distributionally robust optimization under moment uncertainty with application to data-driven problems. *Operations Research,* 58(3): 595–612.

Desiraju R., Moorthy S. 1997. Managing a distribution channel under asymmetric information with performance requirements. *Management Science,* 34(12): 1628–1644.

Eeckhoudt L., Gollier C. Schlesinger H. 1995. The risk averse and prudent newsboy. *Management Science,* 41: 786–794.

Einhorn H.J., Hogarth R.M. 1986. Decision making under ambiguity. *Journal of Business,* 59: 225–250.

Ellsberg D. 1961. Risk, Ambiguity and the Savage Axioms. *Quarterly Journal of Economics,* 75(4): 643–669.

Fox C.R., Tversky A. 1995. Ambiguity aversion and comparative ignorance. *The Quarterly Journal of Economics,* 110(3): 585–603.

Garlappi L., Uppal R., Wang T. 2007. Portfolio selection with parameter and model uncertainty: A multi-Prior approach. *Review of Financial Studies,* 20(1): 41–81.

Gilboa I., Schmeidler D. 1989. Maxmin expected utility with a non-unique prior. *Journal of Mathematical Economics,* 18: 141–153.

Ghirardato P., Maccheroni F., Marinacci M. 2004. Differentiating Ambiguity and Ambiguity Attitude. *Journal of Economic Theory,* 118: 133–173.

Gotoh J., Takano T. 2007. Newsvendor solutions via conditional value-at-risk minimization. *European Journal of Operational Research,* 179(1): 80–96.

Halevy Y. 2007. Ellsberg Revisited: An experimental study. *Econometrica,* 75(2): 503–536.

Heath C., Tversky A. 1991. Preference and belief: Ambiguity and competence in choice under uncertainty. *Journal of Risk and Uncertainty,* 4: 5–28.

Hendricks, K.B. Singhal V.R. 2003. The effect of supply chain glitches on shareholder value. *Journal of Operations Management,* 21: 501–522.

Hsu M., Bhatt M., Adolphs R., Trane D., Camerer C.F. 2005. Neural systems responding to degrees of uncertainty in human decision-making. *Science,* 310(5754): 1680–1683.

Keynes, J.M. 1921. *A Treatise on Probability.* London: Macmillan.

Klibanoff P., Marinacci M. Mukerji S. 2005. A smooth model of decision making under ambiguity. *Econometrica,* 73(6): 1849–1892.

Knight F. 1921. *Risk, Uncertainty & Profit*, Hart, Schaffner, and Marx Prize Essays, no. 31. Boston and New York: Houghton Mifflin.

Kouvelis P., Yu G. 1997. *Robust Discrete Optimization and its Applications.* Kluwer Academic Publishers.

Lau A.H.L., Lau H.S. 2001. Some two-echelon style-goods inventory models with asymmetric market information. *European Journal of Operational Research,* 134: 29–42.

Lurie N.H., Swaminathan J.M. 2009. Is timely information always better? The effect of feedback frequency on performance and knowledge acquisition. *Organizational Behavior and Human Decision Processes,* 108: 315–329.

Natarajan K., Sim M. Uichanco J. 2010a. Tractable robust expected utility and risk models for portfolio optimization, *Mathematical Finance,* forthcoming.

Natarajan K., Sim M. Uichanco J. 2010b. Tractable robust policies for inventory models with demand partitioning. Working Paper.

Natarajan K., Song M., Teo C.P. 2009. Persistency model and its applications in choice modeling. *Management Science,* 55(3): 453–469.

Nesterov Y. 2000. Squared functional systems and optimization problems. In: Frenk H. et al., editors, *High Performance Optimization.* Kluwer Academic Publishers. 405–440.

Perakis G., Roels G. 2008. Regret in the newsvendor model with partial Information. *Operations Research,* 56(1): 188–203.

Popescu I. 2007. Robust mean-covariance solutions for stochastic optimization. *Operations Research,* 55(1): 98–112.

Rockafellar R.T., Uryasev S. 2000. Optimization of conditional value-at-risk. *Journal of Risk,* 2: 21–41.

Savage L.J. 1954. *The Foundation of Statistics.* New York: Wiley.

Scarf H. 1958. A Min-max solution to an inventory problem. In: Arrow K.J., Karlin S., Scarf H., editors, *Studies in the Mathematical Theory of Inventory and Production.* Palo Alto, California: Stanford University Press, pp. 201–209.

Schweitzer M.E., Cachon G.P. 2000. Decision bias in the newsvendor problem with a known demand distribution: Experimental evidence. *Management Science,* 46(3): 404–420.

Shapiro A., Kleywegt A.J. 2002. Minimax Analysis of Stochastic Problems. *Optimization Methods and Software,* 17(3): 523–542.

von Neumann J., Morgenstern O. 1944. *Theory of Games and Economic Behavior,* Princeton, New Jersey: Princeton University Press.

Williams A.C. 1963. A stochastic transportation problem. *Operations Research,* 11(5): 759–770.

Part Two

Integrated Risk Management: Operations and Finance Interface

CHAPTER SIX

Managing Storable Commodity Risks: Role of Inventories and Financial Hedges

PANOS KOUVELIS, RONG LI, AND QING DING

6.1 Introduction

The sourcing, inventory storing, and processing of storable commodities, to be eventually sold in the form of differentiated goods to end-product markets, are cornerstone activities of many business strategies. Examples of storable commodities, also tradable on various exchanges, include oil, LNG (liquid natural gas), steel, precious metals, corn, sugar, DRAM memories, and so on. However, commodity risks can jeopardize even the best thought-out strategies (Tevelson et al. 2007). These days, commodity price risks are even more pronounced and unexpected than before due to shifts in supply-and-demand dynamics and global financial

Handbook of Integrated Risk Management in Global Supply Chains, First Edition. Edited by Panos Kouvelis, Lingxiu Dong, Onur Boyabatli, and Rong Li.

turmoil (Jucker and Carlson 1976). For example, oil marched towards \$150 a barrel in 2008 (even though the current price on June 10, 2010 is about \$76 a barrel, we have seen almost a 169% price increase since January 2009). Steel price soared dramatically to \$1,137 per ton in 2008 and has been swinging between \$335 per ton and \$645 per ton since January 2009. In short, prices of many commodities are now fluctuating as much in a single day as they did in a year in the early 1990s (Wiggins and Blas 2008). Most food commodity prices have doubled from their 2009 lows. Experts predict that late 2010 and 2011 could see a repeat of the commodity price movements of 2008 (see Bryan (2010) for more details and the figures of the food commodity prices in years 2000–2010). The recent wheat price rally in early August 2010 seems to echo this prediction. Due to a severe drought and fire in Russia and a subsequent ban on grain exports from August 15 to December 31, 2010, wheat prices have shown the biggest one-month jump since 1973. European wheat prices jumped 8% on August 2, 2010, the highest in two years; internationally, wheat prices have risen nearly 50% since late June (Blas and Gorst 2010).

For companies that rely on such commodities as production inputs and cannot pass cost increases to their customers, such volatility substantially increases their working capital needs and risks of financial distress. Industries that are close to raw material commodity sources are at the greatest risk. An often cited example is the automotive industry, with exposure to steel and plastic material price risks. As a result, procurement organizations are playing pivotal roles in the financial success of such firms and purchasing managers are expected to have skills never required before. For example, food companies usually allocate procurement activities of commodity inputs to logisticians with limited commodity hedging knowledge and skills. However, as the prices of ingredients that go into corn flakes, chocolate bars, and yogurts squeezed their margins away, food companies are in search of procurement managers with commodity trading skills (Wiggins 2008). Unilever, the multinational food and household products conglomerate, estimates its commodity costs increased in the first half of 2008 to over \$1.5B, the biggest ever annual rise. Hershey, the U.S. chocolate group, saw commodity input costs, such as sugar, peanuts, and cocoa, rise 45% the same year, and is in search to trading skills to implement a \$12M hedging strategy. The same challenges exist in other industries, from mature markets such as automobiles to fast growing markets like high technology products (printers, computers, disk drives, consumer electronics, etc.). Ford posted a loss of more than \$1B on precious metals inventory in the early 2000s due to a misplaced bet on rising prices, and Hewlett-Packard (HP) had a significant risk exposure to flash memory components in the mid-2000s (Nagali et al. 2008). Under significant commodity risk exposure firms are in search of better ways to hedge such exposure in order to lock in supplies, maintain lower costs, minimize earnings volatility, and in the long run, gain a competitive edge.

As argued effectively in Kleindorfer (2008), the growth of commodity exchanges, and derivative instruments defined on them, has offered opportunities, but also research challenges, to integrate traditional forms of bilateral (long-term) contracting with shorter term, market-driven physical and financial transactions for an effective hedging of commodity risks. Access to reasonably liquid spot markets for storable commodities does allow better serving the production needs to fill uncertain customer demands even though at increased exposure to price volatility. The existence of supplier availability risks (i.e., the ability to acquire the right quantity and quality and deliver it at the place and time needed), however, favors the use of long-term contracts with reliable and reasonably proximate suppliers. Thus dual sourcing strategies integrating long-term contracts and short term (spot) market access are a standard storable commodity sourcing practice, with more sophisticated firms using financial hedges as insurance policies to add stability to their input pricing (Knowledge@Wharton July 7, 2008). However, methods for optimally deciding the sourcing allocation between long-term contracts and spot markets, the needed inventory levels of commodity inputs to deal with uncertain end-product demand, and the simultaneous optimal choice of the portfolio of futures (and other derivative) contracts written on commodity exchanges, is a difficult problem, with only limited answers and mostly for non-storable commodities (e.g., electricity) (Kleindorfer 2008). The research presented in this chapter is the first attempt at offering answers to this integrated risk management problem for storable commodities, such as soybeans, metals, and oil.

As Fisher and Kumar (2010) point out, "Too many hedging programs target the nominal risks of 'soiled' businesses rather than a company's net economic exposure-aggregated risk across the broad enterprise that also includes the indirect risks. This soiled approach is a problem, especially in a large multibusiness organizations: managers of business units or divisions focus on their own risks without considering risks and hedging activities elsewhere in the company." In this chapter, we resolve this problem by integrating the financial hedge, which is typically used to hedge the commodity *price risk* (i.e., the price volatility), and the operational hedge (long-term contracts and spot markets), which is typically used to hedge the commodity *consumption volume risk* (i.e., the volatility of consumed commodity volume to meet the uncertain end-product demand). Ignoring the end-product demand uncertainty, most "soiled" businesses aim to mitigate the price risk only. A few, such as the food companies mentioned above, also notice the impact of the consumption volume risk and are looking for procurement personnel with commodity trading experiences. By doing so, these companies are practically linking their financial hedging activities (typically residing in the risk management or finance division of a firm) with their procurement activities (residing in the operations division), often in the absence of an integrated risk management framework and associated support tools. Through this research, we hope to understand how firms can effectively and efficiently manage both commodity risks (price and consumption volume) via integrating sourcing decisions with financial hedging decisions.

6.1.1 PROBLEM SETTING

The fundamental setting of our problem is a firm procuring a commodity input from two sources: a supplier under a long-term contract and a commodity exchange (spot market). The firm processes the commodity inputs into differentiated final products facing uncertain demands in their markets. As argued and shown empirically by many finance researchers, "variability in cash flows now disturbs both investment and financing plans in a way that is costly to the firm" and thus hedging to reduce this variability can increase the firm value (see Froot et al. 1993 and the references therein). The firm aims to hedge the cash flow volatility, driven by both commodity prices and demands, and thus is interested in integrating the trading of futures contracts and other derivative contracts at the commodity exchange with inventory management decisions. For example, Emerson Motor Technologies,[1] headquartered in St Louis and selling a large product line of electro-mechanical motors for various applications, is working with U.S. Steel on purchasing various grades of steel via minimum quantity contracts at an agreed upon long-term price (with typical contract duration of three years). The company is meeting its regional needs (it has factories in the U.S., Mexico, and China) by working with various metal exchanges for spot procurement (mostly with the London Metal Exchange [LME], the Chicago Board of Trade [CBOT] and the China-based Dalian Commodity Exchange [DCE]). These exchanges offer Emerson and other steel consumers derivative products for steel, with futures contracts being the most commonly offered.

In this chapter, we dynamically maximize the total cash flow under mean-variance (MV) criteria to determine time-consistent optimal policies for inventory and financial hedging portfolios. As discussed earlier, the long-term contract price and quantity are typically negotiated for supplier availability risk concerns, which is out of the scope of this research. Thus we assume the long-term contract terms are exogenous and focus on the periodic spot market procurement to meet the uncertain end-product demand. Intuitively, the buyer should build an optimal portfolio of all financial hedging contracts available in the commodity exchange (e.g., futures contracts, call and put options with different maturity times). We argue that we only need to focus on the financial hedging contracts expiring a period after the transaction. Note that futures and forward contracts, although possessing practical differences, are treated as the same analytically (Geman 2005). Unlike forward contracts, typically traded over the counter, futures contracts are tradable in exchanges and thus have better liquidity. Therefore, we consider futures rather than forward contracts. We derive dynamically optimal policies for inventory and financial hedging for various scenarios of available hedging contract choices including: (1) use of a single, and of the same type, hedging contract across all periods; (2) use of the optimal single hedging contract among all available ones in each period; and (3) use of an optimal hedging portfolio open to allocating funds among all the financial contracts (futures, call or put options) available in the commodity exchange. We investigate the interaction between the inventory and the financial hedging and their effects to the buyer's inventory level, mean profit, profit variance, and MV utility.

[1] Information obtained via discussions with Ray Keefe, VP-Manufacturing and Ken Poczekaj, VP-Global Supply Chain of Emerson in May 2010.

6.1.2 OVERVIEW OF RESULTS

We provide a quick preview of our model results:

- We characterize the optimal inventory (base stock) policies for the single and multiple period problems. The base stock levels are dependent on the type of financial hedging used. However, when we know that the firm hedges using futures, alone or together with other hedges, we obtain a myopic base stock policy, which does not require any further details on the structure of the optimal hedging portfolio. The base stock solutions display insightful overage-underage-cost trade-offs, similar to those obtained for the traditional newsvendor model. The overage and underage costs are state-dependent and clearly capture the variance effect and the contribution of financial hedging, and interestingly, reflect whether or not the firm has speculative motivations.
- We derive optimal financial hedging policies for single and multiple period problems, with closed-form hedging quantities. We show that the optimal hedging quantities are heavily dependent on the inventory decision in the current period and all optimal inventory and hedging decisions in future periods. This result emphasizes the need for cross-functional integration for effective commodity risk management, with a particular burden placed on risk managers for understanding the firm's inventory policies.
- In single period settings, consistent with the results in Ding et al. (2007), we show that the futures contract is an optimal hedge and the financial hedges help raise the risk averse buyer's inventory level closer to the risk-neutral optimal inventory level when the risk premium of any financial hedge is zero. Complementing to the existing literature for single period problems, Kouvelis et al. (2010) proves a monotonic result: call (put) options with lower (higher) strike prices perform better (in terms of maximizing the MV utility) and lead to a higher optimal inventory level. When the risk premium is not zero, however, the financial hedges may lower the risk averse buyer's inventory level. For multiperiod settings, the monotonicity results do not hold. Contrary to the findings in Smith and Nau (1995), myopic hedging (with futures contracts) is not optimal in our model. Additionally, unlike in the single period problem, financial hedges may lower the risk averse buyer's inventory level even when their risk premium is zero.
- Finally, we clearly describe the role of the long-term contract, spot market, and financial hedges in dealing with demand volatility of the end-products and price volatility of commodity inputs. Our computational study shows the advantages of integrating physical and financial risk management, with the integrated long-term and short-term contracting delivering the major impact on both mean and variance of the cash flows. The employment of relevant financial hedges allows further control on the variance of the cash flows with moderate benefits (or losses) on mean profits.

The structure of the remainder of the chapter is as follows. In Section 6.2, we review relevant literature and carefully position our work within it. Section 6.3

introduces all relevant notation and important assumptions for our multiperiod model. Section 6.4 formally states the model and provides the optimal inventory and hedging policies for the case of a single (and of the same type) hedge being used across all periods. Section 6.5 deals with the general case, which allows the use of a portfolio of hedges consisting of all possible futures, call and put options. Section 6.6 offers insights on the role and impact of operational and financial hedges on profitability, cash flow variances, and service levels. Two numerical examples of our model application and results are stated in Section 6.7. We conclude with managerial insights and summary of important results in Section 6.8. All proofs can be found in Kouvelis et al. (2010).

6.2 Literature Review

Our work falls under the general themes of "integrated physical and financial risk management in supply chains" and "hedging commodity risks in supply management," which are both expertly reviewed by Kleindorfer (2008 and 2010). For an earlier review on the literature on supply contracting and spot markets, please see Kleindorfer and Wu (2003). The more general field of supply chain contracts is of passing relevance to our work, and we refer the readers to Cachon (2003). In this section, we review in detail the literature most closely related to our paper, and, in particular, the research on integrated long-term and short-term (spot market) contracts.

A general framework with integrated long term-short term contract decisions for mostly non-storable goods is presented in Wu and Kleindorfer (2005). A single-period model is developed to analyze business-to-business (B2B) transactions in supply chains where a buyer and multiple sellers can either contract for delivery in advance (the buyer purchasing "call options" from the sellers) or trade on spot. The authors characterize the structure of the optimal portfolio of contracting with sellers and spot market transactions. For a more extensive review of the related work on non-storable commodities, we refer the readers to the references in Wu and Kleindorfer (2005) and Wu et al. (2002).

For storable commodities, Lee and Whang (2002) is the first to integrate after sales spot market considerations within a newsvendor ordering framework, and thus effectively endogenizes the salvage value used in these models. Peleg et al. (2002) is among the early works on long term–short term integrated sourcing. A stylized two period model is developed to consider a risk-neutral manufacturer who can choose between three alternative procurement strategies: (1) a long-term contract with a single supplier; (2) an online search, in which multiple suppliers are contacted for a price quote; and (3) a combined strategy. The authors characterize conditions under which each of the three alternatives is preferred. In contrast, neither of these works study the specifics of storable commodities (i.e., the availability of buying and selling to spot markets at random prices as well as the availability of the financial derivatives written on commodity prices). Furthermore, these works do not consider issues related to cash flow volatility and hedging for risk management purposes with a risk averse buyer, which are issues of prominence in our work.

Goel and Gutierrez (2006) recently explicitly addresses issues specific to commodity sourcing contexts. The authors analyze a multiperiod procurement problem for a risk-neural manufacturer who procures commodities from spot and futures markets and derive an optimal procurement policy. This work incorporates transaction costs for spot market procurement and endogenizes convenience yield values and their implications for inventory holding costs from the observed spot and futures market prices. Risk aversion concerns and financial hedging of cash flow volatility are, however, not modeled in this work. With rich institutional details of the fed-cattle supply chain, Boyabatli et al. (2011) offers a lucid picture of a beef processor's (e.g., meat packers such as Tyson Foods) problem in these environments via a stylized single period model. The risk neutral processor first contracts for a number of fed-cattle with a feedlot operator, facing demand for beef products and spot price uncertainties. After the uncertainties are resolved, the processor then procures in the fed-cattle spot market, processes under capacity constraints, and then fills demand of two downward substitutable products, program beef and commodity beef. Optimal long term-short term procurement and processing decisions in this proportional production environment are made in the presence of spot market transaction costs, economies of scale in processing, quality differences, and correlated end-product demand. In contrast, our work, although less rich in industry-specific institutional details, is able to handle multiperiod settings, risk aversion, and financial hedging of storable commodities.

There is very limited amount of research on commodity procurement with financial hedging, with most of the existing literature considering non-storable commodities, such as electricity and liquid natural gas (see, e.g., Smith and Nau 1995 and Bodily and Palacios 2007).

Finally, to conclude our review we mention a paper with a methodological similarity to our chapter: Basak and Chabakauri (2010) adopts a dynamic MV utility model, in continuous and discrete time, for an asset allocation problem, where the asset includes a risky stock and a bond. The authors solve for the optimal real-time asset allocation decisions and effectively resolve the well known time-inconsistency issue of the MV criterion (i.e., "optimal decisions" for future times are all determined at the start of the horizon and thus practitioners, observing real-time information, will have incentives to deviate from these decisions) for the first time in the existing literature. Their handling of this issue (via solving a dynamic MV utility model) has been incorporated into our methodology to consider important factors not considered in their model, such as physical inventory, demand uncertainty, and financial hedging.

6.3 Problem Description

The decision maker in our problem is a storable-commodity buyer who processes (or manufactures) this single commodity as an input to make an end-product, which is then sold at a differentiated goods market at an exogenous market price. We first list the sequence of events for each period. At the start of each period, the buyer procures the commodity from two sources: the long-term supply and the spot market. The procurement, together with any on-hand inventory of the commodity, is then processed to meet the uncertain demand. Unmet demand is assumed lost and the excess inventory of the commodity is then carried over to the next period. To account for significant setup time and processing time in each period, we do not allow spot procurement in the middle of the period. It is reasonable to assume that if profitable, the buyer procures primarily for production, with speculation as a secondary purpose. Specifically, we find that if the gross margin (i.e., the revenue less the procurement and processing cost) is positive, the buyer is primarily production-driven (might be speculative at the same time) and processes the commodity to meet the demand. However, if the gross margin is zero or negative, production is not profitable and the buyer only makes speculative purchases in the commodity market. We elaborate on this case in Section 6.4.3.

6.3.1 NOTATION AND ASSUMPTIONS

We next list our notation and assumptions for period n, $n = 1, \ldots, N$. The decision variables are denoted in the last two bullets. We follow the convention of denoting random variables by uppercase letters and their realizations by corresponding lower case letters.

I. $\alpha \in (0, 1)$: the period discount factor for the buyer's cash flow determined as $\frac{1}{1+r}$, where r is the buyer's expected rate of return per period. In finance, r is determined using the Capital Asset Pricing Model (CAPM), which requires the expected return of the market and the risk free rate of return, r_f. Only for presentation convenience, α and r are assumed constant across periods.

II. $\lambda \geq 0$: the absolute risk aversion of the buyer used in mean-variance utility functions ($U = E[\cdot] - \frac{\lambda}{2} V[\cdot]$).

III. $D_n \geq 0$: the buyer's random end-product demand in period n defined on probability space $(\Omega, \mathcal{F}, P)$ with an increasing[2] cumulative distribution function (cdf) $F_n(\cdot)$, where $\mathcal{F}$ is generated by the demand and spot price processes. We assume different periods have mutually independent demands.

IV. $w \geq 0$: the wholesale price of the commodity under the long-term contract with a given fixed quantity $q \geq 0$. Note that when $q = 0$, the long-term contract disappears. Thus single-sourcing from the spot market is a special case of our framework.

V. $S_n \geq 0$: the random spot market price for the commodity input at the start of the period n, defined on $(\Omega, \mathcal{F}, P)$. We assume $\{S_n\}_{1 \leq n \leq N+1}$ is Markovian. Although we allow correlation between S_n and D_n, we assume that given $S_n = s_n$, D_n is independent of $S_{n+1}, \ldots, S_{N+1}$. This assumption is similar to the third assumption for the "partially complete" market in

[2] Increasing cdf is only assumed for presentation convenience. Our analysis and results are applicable to any cdf by, for example, simplify replacing "increasing" with "non-decreasing."

Smith and Nau (1995). This assumption is especially reasonable in our problem as current end-product demand does not influence the future spot market price of the commodity. For example, the demand for air conditioning motors this month at Emerson Motor Technologies generally does not influence the steel spot price next month or after.

VI. $r_n \geq 0$: the unit revenue of the end-product sold in period n, excluding the processing cost.

VII. $h_n \geq 0$: the unit inventory holding cost of the commodity in period n.

VIII. $K_{i,n} > 0$, $\beta_{i,n}$: the strike price and no-arbitrage price (i.e., determined under risk-free probability measure Q) paid upon transaction for hedging contract i, $i = f$ (futures), c (call option), p (put option). A call (put) option is the right, but not obligation, to buy (sell) the commodity at the strike price on the expiration date.

   A. We focus on financial contracts expiring in the next period. This is because: (1) the use of financial contracts with varying expiration dates can be similarly analyzed without any new insights; and (2) financial contracts with later expiration dates (e.g., after two periods) are still up for trading a period later and can be traded then.

   B. $\chi_i(S_{n+1})$: the payoff function for hedging contract i, $i = f, c, p$, where for

      1. futures: $\chi_f(S_{n+1}) = S_{n+1} - K_{f,n}$, where $K_{f,n} = (1 + r_f)E_Q[S_{n+1}]$ and $\beta_{f,n} = 0$.

      2. call option: $\chi_c(S_{n+1}) = (S_{n+1} - K_{c,n})^+ - \beta_{c,n}/\alpha$, where $\beta_{c,n} = \frac{E_Q[(S_{n+1}-K_{c,n})^+]}{1+r_f}$.

      3. put option: $\chi_p(S_{n+1}) = (K_{p,n} - S_{n+1})^+ - \beta_{p,n}/\alpha$, where $\beta_{p,n} = \frac{E_Q[(K_{p,n}-S_{n+1})^+]}{1+r_f}$.

IX. $E[\chi_i(S_{n+1})]$: the buyer's expected payoff for hedge i traded in period n, $i = f, c, p$. If the buyer uses risk-free probability measure and interest rate (i.e., $P = Q$ and $r = r_f$), the expected payoff for any financial hedge is zero and so is the risk premium, which is, for example for call options, defined as $E[(S_{n+1} - K_{c,n})^+]/\beta_{c,n} - (1 + r_f) = \frac{E[\chi_c(S_{n+1})]}{\beta_{c,n}}$ (if $r = r_f$) (Dothan 1990). We refer to this assumption as "zero risk premium." Note that this assumption, also referred to as "fairly priced" in Chod et al. (2010), is commonly used in the relevant literature.

X. $z_n \geq 0$: the commodity inventory level after procurement from both sources (i.e., the inventory level available to fill demand D_n) at the start of period n. Unlike in a stock market, one cannot short in the spot market and thus z_n is assumed to be non-negative.

   A. z_n^{i*}: the optimal commodity inventory level when hedging contract i, $i = f, c, p$, 0 (no hedge), is used.

   B. z_n^*: the optimal commodity inventory level when a portfolio hedge is used.

XI. $y_{i,n} \in \mathbb{R}$: the quantity of financial contract i, $i = f, c$, and p, traded at the start of period n, $n = 0, \ldots, N+1$, where $y_{i,n} < 0$ if contract i is sold and $y_{i,n} > 0$ if contract i is purchased. Without loss of generality, we assume the same unit (e.g., tons of steel) is used for z_n and $y_{i,n}$.

A. $y_{i,n}^{1*}$: the optimal hedging quantity of hedging contract i when a single hedge is used.

B. $y_{i,n}^{*}$: the optimal hedging quantity of hedging contract i when portfolio hedge is adopted.

6.3.2 THE UTILITY FUNCTION

We consider a risk averse buyer of a storable commodity who manages his cash flow volatility by dynamically optimizing a MV utility function of the cash flows with real-time information. Specifically, at the start of each period, the buyer maximizes the MV utility function of the net present value (NPV) of the profit-to-go. The MV analysis of Markowitz, in a single-period framework, is the foundation of modern portfolio theory and has been widely used for measuring the risk aversion in both academia and industry Basak and Chabakauri (2010). It has also inspired the development of multiperiod portfolio choice literature and been adopted in the operations literature. As Van Mieghem (2007) notes, "Similar to financial portfolios, the effect of risk aversion on the configuration of a portfolio of real assets is often illustrated using a MV formulation." It is well known that the MV criterion is consistent with the expected utility criterion when the firm's utility function is quadratic or the cash flow of the firm follows a two-parameter distribution (e.g., Normal distribution) Jucker and Carlson (1976). The two practical benefits of the use of the MV formulation contribute largely to its attractiveness in both academia and industry: (1) easy implementation and (2) providing good recommendations with the efficient frontier even when the firms do not know their utility functions (Van Mieghem 2003).

It is important to note that the MV utility functions, when applied to the NPV of the profit-to-go, are similar to the inter-period utility functions, which are appropriate for characterizing risk aversion Alexander and Sobel (2006). Solving such MV utility functions dynamically is challenging due to the lack of the iterated-expectations property, but has the benefit of resolving the well known time-inconsistency issue with the MV utility. Detailed discussion on this can be found in Basak and Chabakauri (2010), which is believed to be the first paper to resolve the time-inconsistency issue in the portfolio choice literature. In contrast, although our problem is not an asset allocation problem, it does bear methodological similarities with the inventory and financial hedges playing equivalent roles to the risky stocks. Moreover, our problem is substantially more complex with the consideration of physical inventory, demand uncertainty, and financial hedging.

In conclusion, at the start of period n, $n = 1, \ldots, N$, the buyer optimizes the inventory level, z_n, and the amount of the financial contract i, $y_{i,n}$, $i = f, c, p$. At the end of the horizon, we assume that the buyer receives no supply from the supplier and only trades in the spot market. Having no demand to fill, the buyer should simply sell all the excess inventory to the spot market.

6.4 Optimal Policy for Single Contract Financial Hedging

We first study the inventory and hedging policies for the buyer who chooses to employ a single hedge, either futures, call or put option. The same type of hedge is used across all periods. Note that futures and options are the most commonly used hedging contracts in practice. Our analysis with a single hedge highlights important aspects of a risk averse buyer's behavior on balancing quantity risk and price risk through the integrated use of operational and financial hedges.

6.4.1 THE BUYER'S UTILITY FUNCTION WITH SINGLE CONTRACT FINANCIAL HEDGING

At the start of period n, observing the real-time information, the buyer needs to make the inventory decision, $z_n \geq 0$, and the hedging decision, $y_{i,n} \in \mathbb{R}$, $i = f, c, p$. He then processes the commodity to meet the uncertain end-product demand if profitable.

Note that r_n represents the unit revenue excluding the processing cost. The relevant buyer's gross margin is $r_n - s_n$, based on which the buyer decides whether to produce to fill his customer demand. Note that the purchasing cost paid to the long-term supplier each period is a sunk cost. When the gross margin $r_n - s_n$ is greater than 0, the buyer should procure for processing to meet the demand. Let $\tilde{\pi}^i_n(s_n, x_n, y_{i,n-1}, z_n)$ denote the buyer's profit function in period n (if the buyer restricts himself to the use of a single hedge of type i, $i = f, c, p$, across all periods), for any given spot price s_n, on-hand inventory level of the commodity x_n, quantity of contract i traded in the previous period $y_{i,n-1}$, and final inventory level of the commodity $z_n \geq 0$. Thus, we have

$$\begin{aligned}
&\tilde{\pi}^i_n(s_n, x_n, y_{i,n-1}, z_n) \\
&= y_{i,n-1}\chi_i(s_n) - wq - s_n[z_n - (x_n + q)] + r_n(z_n \wedge D_n) - h_n(z_n - D_n)^+ \\
&= s_n x_n + y_{i,n-1}\chi_i(s_n) + (s_n - w)q + (r_n - s_n)(z_n \wedge D_n) \\
&\quad -(s_n + h_n)(z_n - D_n)^+ \qquad (6.1)
\end{aligned}$$

where $a \wedge b = \min\{a, b\}$ and $a \vee b = \max\{a, b\}$ for any $a, b \in \mathbb{R}$.

As the second equation shows, under the assumption of zero bid-ask spread for the spot market, the profit can be assessed as coming from (1) selling the on-hand inventory of the commodity at s_n per unit, (2) exercising the hedging contract, (3) selling the procurement from the supplier directly to the spot market, and (4) purchasing z_n units of the commodity from the spot market and processing them to meet the demand, minus the holding cost of the excess commodity. This observation helps characterize the optimal policies and investigate the impact of each individual payoff on the total utility.

The excess commodity, at amount $(z_n - D_n)^+$ (i.e., the on-hand inventory of the commodity at the start of period $n+1$), will be sold at price S_{n+1}. By shifting the payoff of this transaction to period n, we see that the buyer always starts each period with zero on-hand inventory. We now remove the state variable for the on-hand inventory level and re-define the buyer's profit function in period n, denoted by $\pi_n^i(s_n, y_{i,n-1}, z_n)$, as follows.

$$\pi_n^i(s_n, y_{i,n-1}, z_n) = y_{i,n-1}\chi_i(s_n) + (s_n - w)q + (r_n - s_n)(z_n \wedge D_n) \\ +(\alpha S_{n+1} - s_n - h_n)(z_n - D_n)^+. \quad (6.2)$$

When the gross margin $r_n - s_n \leq 0$, since production is not profitable, the buyer makes speculative purchases of the commodity. In this case, the buyer's profit function becomes:

$$\pi_n^i(s_n, y_{i,n-1}, z_n) = y_{i,n-1}\chi_i(s_n) + (s_n - w)q + (\alpha S_{n+1} - s_n - h_n)z_n. \quad (6.3)$$

At the end of the horizon, facing no customer demand, the buyer simply sells all the excess commodity to the spot market (i.e., $z_{N+1} \equiv 0$). Thus, regardless of the sign of $r_n - s_n$, we have

$$\pi_{N+1}^i(s_{N+1}, y_{i,N}, z_{N+1} \equiv 0) = y_{i,N}\chi_i(s_{N+1}) \quad (6.4)$$

Using π_n^i to represent the profit in period n, we formally define the buyer's MV utility function for period n, $n = 1, \ldots, N+1$, by

$$U_n(z_n, y_{i,n}|s_n, y_{i,n-1}) = E\left[\sum_{k=n}^{N+1} \alpha^{k-n}\pi_k^i\right] - \frac{\lambda}{2}V\left[\sum_{k=n}^{N+1} \alpha^{k-n}\pi_k^i\right], \quad (6.5)$$

where $\pi_{n+1}^i = \pi_{n+1}^i(S_{n+1}, y_{i,n}, Z_{n+1}^{i*})$ and $\pi_k^i = \pi_k^i(S_k, Y_{i,k-1}^{1*}, Z_k^{i*})$, $k = n+2, \ldots, N+1$, in which Z_k^{i*} and $Y_{i,k}^{1*}$ are random variables representing the optimal decisions for period k, $k \geq n+1$ (randomness coming from spot prices $S_{n+1}, \ldots, S_k$ and demands $D_n, \ldots, D_{k-1}$ which are not observable at the decision time, the start of period n).

At the end of horizon, facing no risks, the buyer should not consider hedging, i.e., $y_{i,N+1}^{1*} = 0$, and thus the buyer's utility function can be simplified using (6.4) as

$$U_{N+1}(z_{N+1}^{i*} = 0, y_{i,N+1}^{1*} = 0|s_{N+1}, y_{i,N}) = y_{i,N}\chi_i(s_{N+1}). \quad (6.6)$$

If the demand is constant (or perfectly correlated with the spot price), financial contracts can be used to perfectly hedge the spot price risk. But when the demand in random, but not perfectly correlated with the spot price, this is not possible anymore. For example, let us look at the hedging decision for the last period, where D_N and S_{N+1} are independent. Since we gain profit from selling the excess commodity to the spot market, we should sell futures contracts to counterbalance lower spot prices. These futures contracts, obviously, cannot hedge the demand risk.

The hedging decisions for any other period are more complex. For illustration simplicity, let us suppose the demand is independent of the spot price and only financial contracts with zero risk premium are considered. At the start of period n, we hedge (via trading financial contracts contingent on the spot price S_{n+1}) the volatility of the total cash flow earned from period n onwards. As the period profit definition, given by (6.2), indicates, the cash flow affected by S_{n+1} includes: (1) $S_{n+1}(z_n - D_n)^+$, (2) $(S_{n+1} - w)q$, (3) $-S_{n+1}Z_{n+1}^{i*}$, (4) $r_n(Z_{n+1}^{i*} \wedge D_{n+1})$, and (5) $-h_{n+1}(Z_{n+1}^{i*} - D_{n+1})^+$. Let us consider the commodity risks involved in each term in order to better understand what we are hedging. Terms (1) and (2) are profits exposed to the risks of lower spot prices. Term (3) is the procurement cost and thus is exposed to the risk of higher spot prices if the inventory level Z_{n+1}^{i*} were constant. Term (4) is revenue subject to having enough inventory. Term (5) is the holding cost and thus is exposed to the risk of too much inventory. Note that the inventory level Z_{n+1}^{i*} is driven by both the demand D_{n+1} and the spot price S_{n+1} (as cost), and higher spot prices result in a lower inventory level. Therefore, terms (4) and (5) are also exposed to spot price risks. In summary, we should hedge against lower spot prices for terms (1), (2), and (5), while we should hedge against higher spot prices for term (4). For term (3), however, the direction of the hedge is unclear.

6.4.2 OPTIMAL POLICY FOR EACH PERIOD

We characterize the optimal inventory and hedging policy for period n in this section. Although determined simultaneously, the inventory and hedging decisions for each period have one period difference for their effective times, due to the one period time lag between the transaction and the exercise of the hedging contracts.

For readers' convenience, we start by simplifying the notation. First, we denote *the future profit* for period n, $n < N$, for the case of using contract i, $i = f, c, p$, as the single hedge by

$$\Pi_n^i(\mathbf{S}) = \pi_{n+1}^i(S_{n+1}, y_{i,n}, Z_{n+1}^{i*}) - y_{i,n}^{1*}\chi_i(S_{n+1}) + \sum_{k=n+2}^{N+1} \alpha^{k-n-1}\pi_k^i(S_k, Y_{i,k-1}, Z_k^{i*}) \tag{6.7}$$

where $\mathbf{S} = (S_{n+1}, \ldots, S_{N+1})$.

For the last period, $\Pi_N^i(\mathbf{S}) \equiv 0$. Second, we omit the condition $S_n = s_n$ in the conditional expectation terms in the utility functions, such as $E[\alpha S_{n+1}(z_n - D_N)^+|S_n = s_n]$. Third, since the first order condition for the inventory $\frac{dU_n(\cdot)}{dz_n} = c_u(n, z_n)\bar{F}_n(z_n) - c_o(n, z_n)F_n(z_n) = 0$ displays similarities as the standard newsvendor solution, we define $c_o(n, z_n)$ and $c_u(n, z_n)$ as the overage and underage costs, respectively, where:

$$c_o(n, z_n) = c_{o1}(n) + c_{o2}(n, z_n) \text{ and } c_u(n, z_n) = c_{u1}(n) + c_{u2}(n, z_n) \tag{6.8}$$

Where $i = f, c, p$:

$$\begin{aligned}
c_{o1}(n) &= s_n + h_n - E[\alpha S_{n+1}] + \lambda\alpha^2 Cov(S_{n+1}, \Pi_n^i(\mathbf{S})) - c_{o1}(n, \chi_i(S_{n+1})) \\
c_{o2}(n, z_n) &= \lambda\alpha^2 V[S_{n+1}]E[(z_n - D_n)^+] - c_{o2}(n, z_n, \chi_i(S_{n+1})) \\
c_{u1}(n) &= r_n - s_n \\
c_{u2}(n, z_n) &= -\lambda \left((r_n + h_n - E[\alpha S_{n+1}])^2 + \alpha^2 V[S_{n+1}]\right) E[(z_n - D_n)^+] \quad (6.9)
\end{aligned}$$

The hedging related terms are:

$$\begin{aligned}
c_{o1}(n, \chi_i(S_{n+1})) &= \frac{\alpha Cov(S_{n+1}, \chi_i(S_{n+1}))}{V[\chi_i(S_{n+1})]} \\
&\quad \times \left(\lambda\alpha Cov(\chi_i(S_{n+1}), \Pi_n^i(\mathbf{S})) - E[\chi_i(S_{n+1})]\right) \\
c_{o2}(n, z_n, \chi_i(S_{n+1})) &= \lambda\alpha^2 \frac{Cov^2(S_{n+1}, \chi_i(S_{n+1}))}{V[\chi_i(S_{n+1})]} E[(z_n - D_n)^+] \qquad (6.10)
\end{aligned}$$

In contrast to the standard newsvendor solution, the overage or the underage cost in our problem consists of two parts, a constant cost ($c_{o1}(n)$ or $c_{u1}(n)$) and a monotonic variable cost ($c_{o2}(n, z_n)$ or $c_{u2}(n, z_n)$). Indeed, $c_{o2}(n, z_n)$ increases in z_n (except when the futures contract is used and thus $c_{o2}(n, z_n) = 0$), while $c_{u2}(n, z_n)$ decreases in z_n. More interpretation of these costs is provided later in this section. We show that a base-stock policy is optimal for every period.

Proposition 6.1 *Given spot price s_n, $n \le N$, for $i = f, c, p$, the optimal quantities are:*
Hedging:

$$\begin{aligned}
y_{i,n}^{1*} = &- \left\{E[(z_n^{i*} - D_n)^+]1_{\{r_n > s_n\}} + z_n^{i*}1_{\{r_n \le s_n\}}\right\} \frac{Cov\big(S_{n+1}, \chi_i(S_{n+1})\big)}{V[\chi_i(S_{n+1})]} \\
&- \frac{Cov(\chi_i(S_{n+1}), \Pi_n^i(\boldsymbol{S}))}{V[\chi_i(S_{n+1})]} + \frac{E[\chi_i(S_{n+1})]}{\lambda\alpha V[\chi_i(S_{n+1})]}
\end{aligned}$$

If $E[\chi_i(S_{N+1})] = 0$ (zero risk premium), we have $|y_{f,N}^{1*}| \le |y_{j,N}^{1*}|$, $j = c, p$.
Inventory:
When $r_n - s_n > 0$ and $c_{o1}(n) > 0$, $z_n^{i*} > 0$ is the unique solution to the first order condition (FOC), that is, $c_o(n, z_n^{i*})F_n(z_n^{i*}) = c_u(n, z_n^{i*})\bar{F}_n(z_n^{i*})$.

We start by interpreting the hedging result. As discussed in the previous section for the last period, the buyer hedges the profit earned from selling the excess commodity. For zero risk premium, the hedging contributes to profit variance reduction only. Thus, it is effective via a short position by selling a futures contract or a call option or buying a put option. In this case, the hedging amount $y_{i,N}^{1*}$, given by the formula above, includes only the first term. Specifically, for each unit of excess commodity to sell, the absolute hedging amount is $\left|\frac{Cov(S_{N+1}, \chi_i(S_{N+1}))}{V[\chi_i(S_{n+1})]}\right|$ (≥ 1 and $= 1$ when $i = f$). Furthermore, in Section 6.1, we show that the futures contract is indeed the best single hedge for the last period. Note that the hedging contracts are similar to return contracts, but with fixed return quantities. The difference between this fixed quantity and the excess commodity will be cleared by trading in the spot market.

For any other period, however, we should also hedge against the risk in the remainder of the cash flow (i.e., the future profit $\Pi_n^i(\mathbf{S})$). When nonzero risk premium is considered, hedging also contributes to the mean profit. Indeed, the hedging quantity is determined by balancing the two sources of contributions to the utility, the profit variance reduction and the mean profit gain. In summary, the optimal hedging quantity consists of three parts, each of which has the following function: (1) hedge the risk in the current period profit $\alpha S_{n+1}(z_n - D_n)^+$ (or reduce the corresponding profit variance) by holding a short position, (2) hedge the risk in the future profit $\alpha \Pi_n^i(\mathbf{S})$, and (3) boost the hedging payoff (or increase the mean profit). Therefore, the hedging decision only interacts with the inventory decision for the same period via the term $\alpha S_{n+1}(z_n - D_n)^+$. In other words, the financial hedging only contributes to the inventory management for the current period by partially reducing the utility loss due to overstocking. If we were to hedge only the cash flow in the current period (a myopic hedging), we should use the futures contract (same as we did for the last period).

As previously discussed, in the use of zero bid-ask spread in the spot market, the buyer starts each period with zero inventory. Thus the inventory decision in the current period has no impact on the future decisions. Conversely, if the utility function includes only the mean profit, the future decisions should not influence the current inventory decision. This argument, however, no longer holds when variance is included in the utility function. The current inventory decision and the future decisions are linked via the term $\alpha S_{n+1}(z_n - D_n)^+$. Indeed, their dependence is captured by the covariance between the current period profit and the future profit $\alpha^2 Cov[S_{n+1}(z_n - D_n)^+, \pi_n^i(\mathbf{S})]$. Thus, in general, the optimal policy is not myopic and requires the knowledge of the optimal decisions for all future periods. It is important to note that since the futures contract perfectly hedges the covariance, the optimal inventory policy is therefore myopic. More discussion of this myopic inventory policy is provided in the following section.

Note that due to the use of the multiperiod MV utility and the incorporation of financial hedging, our inventory solution is more complex than the standard newsvendor solution. It, however, reveals similar (but new) insights on the trade-off between the overage ($c_o(n, z_n)$) and underage ($c_u(n, z_n)$) costs. As the corresponding expression shows, the underage cost is the profit loss $r_n - s_n$ less the utility gain from reducing the profit variance (as less inventory leads to a smaller variance). Note that the financial hedge only affects the overage cost since it is effective only when excess commodity exists and is sold to the spot market. Indeed, for the last period, if zero risk premium is assumed, any financial hedge reduces the overage cost, resulting in an inventory increase (as shown in Proposition 6.4). With nonzero risk premium, the financial hedge may lead to less inventory. For any other period, in contrast, the financial hedge reduces the increasing rate of the overage cost $\frac{d}{dz_n} c_{o2}(n, z_n)$, although it may not reduces the overage cost itself (as the future profit also requires hedging). Specifically, the overage cost equals the procurement and holding cost, $s_n + h_n$, less the expected revenue $E[\alpha S_{n+1}]$ (as being sold at S_{n+1}), plus the utility loss (or the profit variance increase) reduced by financial hedging (detailed reduction is shown in (6.10), and plus the "cost" of hedging (or the mean profit decrease, which equals the unit hedging cost, $E[\chi_i(S_{n+1})]$, times the hedging amount, $\frac{Cov(S_{N+1},\chi_i(S_{N+1}))}{V[\chi_i(S_{n+1})]}$). Finally, when the buyer is risk neutral and the financial hedge has zero risk premium (i.e., $\lambda = 0$ and $E[\chi_i(S_{n+1})] = 0$, $i = f, c, p$), the optimal inventory solution reduces to a standard newsvendor solution (i.e., $c_o(n, z_n) = s_n + h_n - E[\alpha S_{n+1}]$ and $c_u(n, z_n) = r_n - s_n$).

6.5 Optimal Policy for a Portfolio of Financial Hedges

We now consider the case that the commodity buyers have access to a portfolio of all available financial hedging contracts in the market. We note that it is sufficient to study hedge portfolios consisting of the futures contracts, call and put options only (see Carr and Madan (2001)).

We first define some additional parameters and decision variables. Let $K_{c,n,i}$ and $\chi_{c,i}(S_{n+1})$, $i = 1, \ldots, n_c$, $K_{p,n,j}$ and $\chi_{p,j}(S_{n+1})$, $j = 1, \ldots, n_p$, denote the strike price and payoff function for the call and put options available at the start of period n, respectively. Without loss of generality, we assume these financial hedges are not replicating each other. Let $\mathbf{y}_n = [y_{f,n}, y_{c,n,1}, \cdots, y_{c,n,n_c}, y_{p,n,1}, \cdots, y_{p,n,n_p}]$ denote the corresponding hedging quantity array. Let $\pi_n(\cdot)$, $\Pi_n(\mathbf{S})$, and $U_n(z_n, \mathbf{y}_n|s_n, \mathbf{y}_{n-1})$[3] denote the buyer's current period profit, total future profit, and MV utility function for period n given that the hedging portfolio is $\mathbf{y}_n$, respectively.

Since the Hessian matrix of the utility function is $H = -\lambda\alpha^2\Sigma$, where Σ is the covariance matrix for random variables S_{n+1}, $(S_{n+1} - K_{c,n,1})^+, \ldots, (S_{n+1} - K_{c,n,n_c})^+$, $(S_{n+1} - K_{p,n,1})^-, \ldots, (S_{n+1} - K_{p,n,n_p})^-$, we know that H is negative semi-definite and thus the utility function is concave. The concavity leads to the optimality of the base-stock inventory policy. In addition, we know that Σ is invertible as the hedges are not replicating each other. For presentation convenience, we let $\Psi_n = [\psi_{f,n}, \psi_{c,n,1}, \cdots, \psi_{c,n,n_c}, \psi_{p,n,1}, \cdots, \psi_{p,n,n_p}]$, where for example $\psi_{c,n,i} = \frac{E[\chi_{c,i}(S_{n+1})]}{\lambda\alpha} - E(z_n^* - D_n)^+ Cov(\chi_{c,i}(S_{n+1}), S_{n+1}) - Cov(\chi_{c,i}(S_{n+1}), \Pi_n(\mathbf{S}))$.

Proposition 6.2 *Given spot price s_n, $n \leq N$, the optimal quantities are:*
Hedging:

$$\boldsymbol{y}_n^* = \Sigma^{-1}\Psi_n. \text{ If } \frac{E[\chi_f(S_{n+1})]}{V[S_{n+1}]} = \frac{E[\chi_{c,i}(S_{n+1})]}{Cov(\chi_{c,i}(S_{n+1}), S_{n+1})} = \frac{E[\chi_{p,j}(S_{n+1})]}{Cov(\chi_{p,j}(S_{n+1}), S_{n+1})}$$

[3] They are similarly defined as $\pi_n^i(\cdot)$ (defined by [6.2]), $\Pi_n^i(S)$ (defined by [6.7]), and $U_n(z_n, y_{i,n}|s_n, y_{i,n-1})$ (defined by [6.5]), respectively, except that the hedging payoff here is the sum of the payoffs from each hedging contract.

For all i and j, which is immediately satisfied given zero risk premium, we have $\boldsymbol{y}_N^* = \left[-E[(z_N^* - D_N)^+] + \frac{E[\chi_f(S_{N+1})]}{\lambda\alpha V[S_{N+1}]}, 0, \ldots, 0\right]$.
Inventory: $z_n^* \equiv z_n^{f*}$ *and is the unique solution to*

$$\begin{aligned}&\left(s_n + h_n - E[\alpha S_{n+1}] + \alpha E[\chi_f(S_{n+1})]\right) F_n(z_n^*)\\ &= \left(r_n - s_n - \lambda\left((r_n + h_n - E[\alpha S_{n+1}])^2 + \alpha^2 V[S_{n+1}]\right) E[(z_n^* - D_n)^+]\right) \bar{F}_n(z_n^*)\end{aligned}$$

This proposition implies that for a single period case, the buyer's optimal hedging portfolio contains futures only if all hedges have zero risk premium, or more generally, if all hedges have an equal "discounted" risk premium, $\frac{E[\chi(S_{N+1})]}{Cov(\chi(S_{N+1}),S_{N+1})}$. For a multiperiod case, however, the buyer's optimal hedging portfolio is comprised of the futures contract, call options with strike prices lower than the futures' price, and all put options with strike prices higher than the futures' price. Similarly as in the single hedge case discussed in the previous section, the optimal hedging quantity of each contract in the portfolio also consists of three parts, which are used to hedge the current period profit, hedge the future profit, and boost the hedging payoff, respectively.

It is important to note that the optimal inventory level, z_n^*, when using the hedging portfolio is myopic and identical to the optimal inventory level, z_n^{f*}, when using futures alone. Intuitively, as long as the futures contract is utilized, the covariance between the current period profit and the future profit is perfectly hedged and thus the inventory policy is myopic and easy to compute. As a result, the financial hedging completely reduces the utility loss in the overage cost, discussed in the previous section. Thus the overage cost $c_o(n, z_n)$ is a positive constant. Since in this case the overage cost is always positive, the buyer has no speculative motives—he procures for production only when the gross margin is positive. Specifically, $z_n^* = z_n^{f*}$ is because the covariance between the hedging payoff and the to-be-hedged payoff $S_{n+1}(z_n - D_n)^+$ for these two cases are equal[4]. However, the hedging payoff also affects other parts in the utility function and therefore using the futures contract alone is not optimal for any period besides the last. Furthermore, although the risk premium of any financial contract used influences its optimal quantity and the utility, z_n^* is only affected by the risk premium of the futures contract.

6.6 Role of the Operational and Financial Hedges

We distinguish the cash flow hedging tools for the risk averse buyer into two categories, the operational hedge and the financial hedge. The operational hedge (i.e., the physical inventory) includes the long-term supply of fixed quantity and the spot market procurement (buying or selling). They are effectively used to

[4] It is because $y_{f,n}^* V[S_{n+1}] + \sum_{i=1}^{n_c} y_{c,n,i}^* Cov(S_{c,i}^+, S) + \sum_{j=1}^{n_p} y_{p,n,j}^* Cov(S_{p,j}^-, S) = y_{f,n}^{1*} V[S_{n+1}]$.

deal with the supply-demand mismatch risk. By its nature, the long-term supply protects the buyer against the commodity price risk, while the spot market protects the buyer against the commodity consumption risk (or the end-product demand risk). Since the procurement cost of the long-term supply is a sunk cost, the effective operational hedge is indeed the spot market procurement. Different financial hedges we study in this paper include single and multiple hedges, comprised of futures, call and put options. They hedge directly the commodity price risk and only indirectly the demand risk (via hedging amounts and correlation of demands and spot prices, and resulting in inventory decisions with service level improvement). As different financial hedges are compared, we assume zero risk premium to facilitate a "fair" comparison.

We next examine how the use of different financial hedging influences the effectiveness of the operational hedge, which is measured by the service level. We then discuss how the use of the operational hedge (inventory) and different popular financial hedges affect the buyer's financial performance, characterized by the mean, variance, and utility of the cash flows. Specifically, we explore the similarity and distinction between the two types of hedges by comparing some different scenarios, which include SS (single sourcing), DS (dual sourcing), DS + i (dual sourcing with a single hedge i, i = f (futures), c (call), p (put)), and DS + HP (dual soucing with a hedging portfolio).

6.6.1 IMPACT ON THE SERVICE LEVEL

Let z_n^{0*} denote the buyer's optimal inventory level period n, $n = 1, \ldots, N$, for the DS case[5]. We have shown that z_n^* (for DS + HP case) is the same as z_n^{f*} (for DS + f case). Therefore, in addition to illustrating how the risk aversion and different financial hedges affect the inventory level (and thus the service level), we present detailed sensitivity results for the optimal inventory level z_n^{f*} as its significance goes beyond the use of futures as a single hedge. We also contrast our results to the relevant literature on operations with financial hedging (exclusive of the commodity market).

Proposition 6.3 *For period n, $n \le N$, we observe:*
(1) For risk-neutral buyers (i.e., when $\lambda = 0$):

- *if* $E[\chi_i(S_{n+1})] = 0$, $z_n^{i*} = F_n^{-1}\left(\frac{r_n - s_n}{r_n + h_n - E[\alpha S_{n+1}]}\right) \triangleq \bar{z}_n^*$, $i = f, c, p$
- *if* $E[\chi_i(S_{n+1})] \neq 0$, z_n^{i*} *is arbitrary and* $|y_{i,n}^{1*}| = \infty$

(2) Sensitivity to the risk aversion (λ): z_n^* is decreasing in λ.
(3) z_n^* is a decreasing function of h_n and $V[S_{n+1}]$.
(4) Service level comparison: $\bar{z}_N^* \ge z_N^* \ge z_N^{c*}, z_N^{p*} > z_N^{0*}$.

---
[5] z_n^{0*} satisfies the FOC with the overage and underage costs modified by omitting the hedging related terms in (6.9).

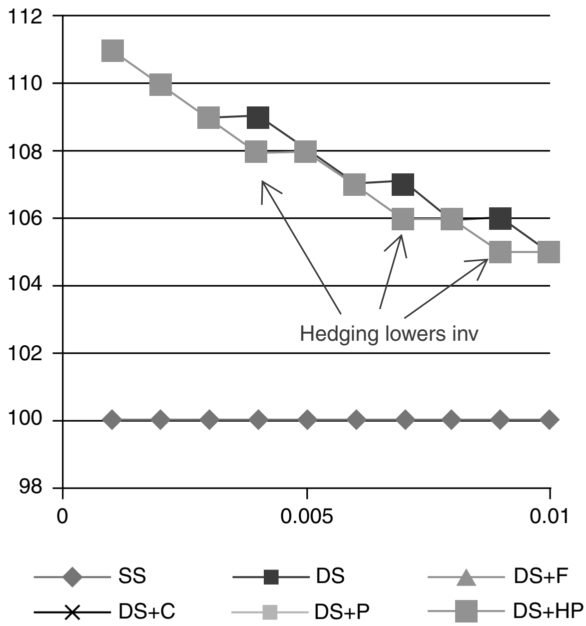


FIGURE 6.1 **Optimal inventory level against $\frac{\lambda}{2} = 0.001, \ldots, 0.01$ at $\rho = -0.5$.**

Intuitively, a more risk averse buyer usually keeps a lower inventory level despite the choice of financial hedging. That is, a risk neutral buyer carries the highest inventory level. Our numerical study (refer to the end of Section 6.7 for details) indicates that when λ is as low as 0.002, risk averse buyers with any financial hedging carry the same inventory amount as risk-neutral buyers (see Figures 6.1 and 6.2). Given zero risk premium, a risk neutral buyer has no interest in financial hedging. In the cases of nonzero risk premium, however, a risk neutral buyer trades pathologically in the derivative market. When futures contracts are used as a single hedge or in a portfolio of hedges, as the inventory holding cost rises, a risk averse buyer lowers his inventory level to increase the mean and decrease the variance of the cash flows. Facing a higher spot price volatility, a risk averse buyer keeps a lower inventory level to reduce the variance. In addition, the higher the risk premium of the futures (more costly hedging), the lower the risk averse buyer's inventory level.

For the last period, when the risk premium for any financial hedge is zero, our results are consistent with the results for the single-period financial hedging problems studied by, among others, Ding et al. (2007). In particular, the risk averse buyer stocks less than a risk-neutral buyer; the use of financial hedging helps raise the risk averse buyer's inventory level. In addition, we show that a better hedge (e.g., a call with a lower strike price or a put with a higher strike price) leads to a higher inventory level due to the better control of the variance of the cash flows. When the risk premium is not zero, however, a different phenomenon is

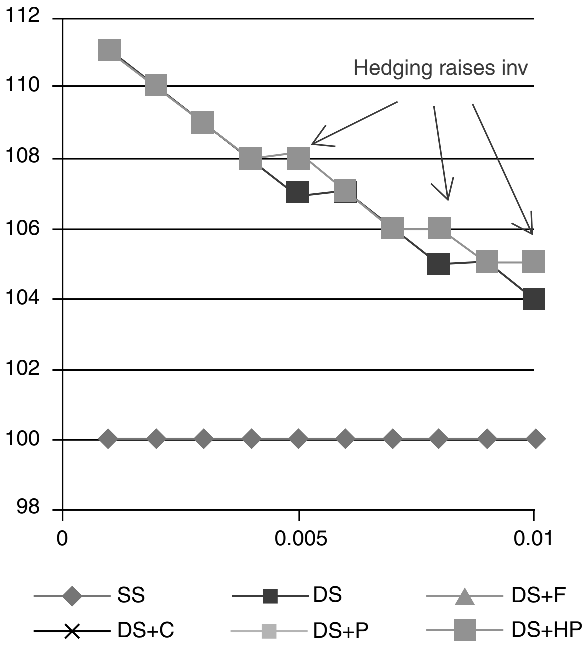


FIGURE 6.2 **Optimal inventory level against** $\frac{\lambda}{2} = 0.001, \ldots, 0.01$ **at** $\rho = 0.5$.

observed: the use of financial hedging, even though it raises the utility value with certainty, may or may not drive up the risk averse buyer's inventory level.

For any other period, however, even when the risk premium for any financial hedge is zero, our result differs from the results for the multiperiod financial hedging problem studied by Smith and Nau (1995). This work adopts intra-period utility functions, displaying the "additive independence" property, to show the optimality of myopic hedging (i.e., hedging the current period cash flow only), under a number of assumptions (e.g., the "partially complete" market), for decomposing the cash flow into its market and private components. In contrast, we use inter-period utility functions, not exhibiting the "additive independence" property, and allow interdependence between the spot price and the customer demand (which implies the cash flow decomposition is infeasible). We characterize a non-myopic optimal hedging portfolio, which is easy to compute and implement. We find that although the use of futures alone is an optimal myopic hedging strategy, it may not be the true optimal. In addition, we observe that the use of financial hedging may lower the risk averse buyer's inventory level (shown in Figure 6.1); the nonzero risk premium further lowers the inventory level. As the financial hedge is also used to hedge the future profit, there no longer exists the nice monotonic relationship between the ranking of the financial hedge and the corresponding inventory level. We observe in the numerical study that a better hedge may correspond to a lower inventory level. In addition, comparing Figures 6.1 and 6.2 we find that the way, in which financial hedging affects the inventory level, may depend on the correlation between demands and spot prices.

6.6.2 IMPACT ON THE MEAN, VARIANCE AND UTILITY

We now discuss how the use of inventory and different financial hedges affect the buyer's financial performance characteristics including the mean, variance, and utility of the cash flows.

Proposition 6.4 *For period n, $n = 1, \ldots, N$, we have the following results.*
(1) Ranking by the utility value (from best to worst): DS + HP, DS + i, DS, and SS (where the comparison of the last two is in expectation).
(2) When $z_n \leq \bar{z}_n^*$ (the feasible region for the risk averse buyer), as z_n increases,

- $E\left[\sum_{k=n} \alpha^{k-n}\pi_k\right]$ *and* $V\left[\sum_{k=n} \alpha^{k-n}\pi_k\right]$ *both increase if the futures contract is used.*
- $E[\pi_N]$ *and* $V[\pi_N]$ *both increase if any or no financial hedge is used.*

We find that, in general, it is better to make use of multiple hedging tools. The operational hedge is more effective (in terms of improving profitability) than the financial hedge if only one is allowed for use because the benefits of making real-time decision is salient. In addition, we use the numerical study to illustrate effectiveness of the inventory and financial hedging in reducing the profit variance. As shown in Figures 6.3 and 6.4, for the case of positively correlated demand and spot price ($\rho = 0.5$), the profit variance reduction by switching from SS to DS is between 82% and 84%; the further profit variance reduction by adding financial hedging is between 14% and 18%. For most cases, a higher inventory level corresponds to a higher mean profit, but it also has a higher profit variance. Thus, an optimal inventory level is chosen to balance the mean and variance of

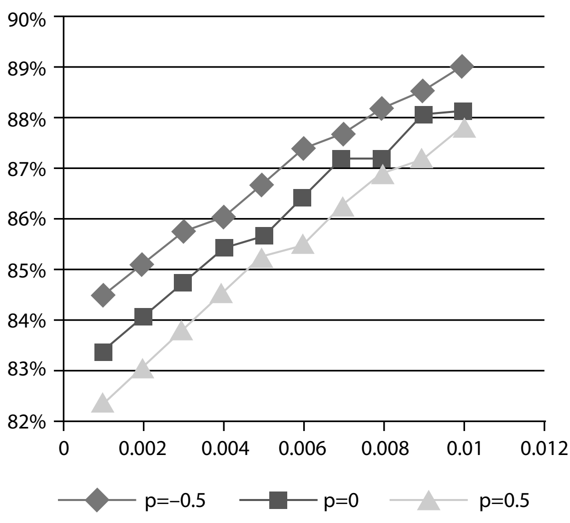


FIGURE 6.3 **Variance reduction from SS to DS against $\frac{\lambda}{2} = 0.001, \ldots, 0.01$.**

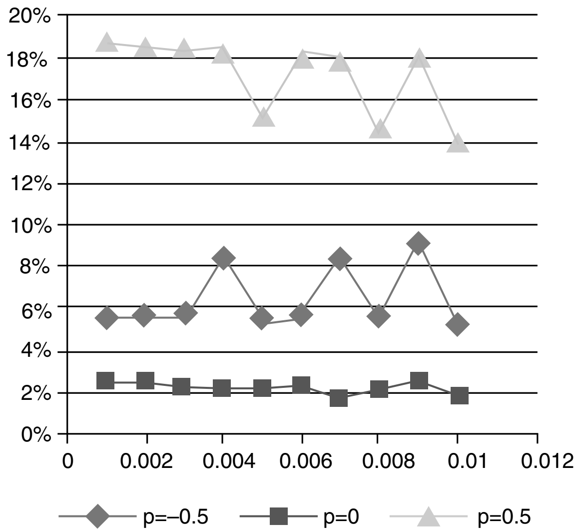


FIGURE 6.4 **Variance reduction from DS to DS + HP against** $\frac{\lambda}{2} = 0.001, \ldots, 0.01$.

the cash flows. As the buyer becomes more risk averse, the profit variance plays a more important role and thus the inventory level is reduced accordingly.

We next use the efficient frontier (see Figures 6.5 and 6.6) to compare the impacts of different hedging strategies for buyers with different risk aversions. Since SS has a significant lower mean but higher variance than all other cases, it is removed from the figures for a clearer view of the other cases. For example, for the case of $\rho = 0.5$, by switching from SS to DS the mean increase is about 7.5% and the variance reduction is above 80%. As his risk aversion increases, the buyer moves to the optimal position with a lower mean and a lower variance for any case. We observe that DS and DS + HP are at the top and the bottom of the graph area, respectively. The other curves for DS + i, $i = f, c, p$, are sitting in the middle with alternating positions as the risk aversion increases. This clearly implies that it is always beneficial to adopt financial hedges, and the choice of a good hedge does have impact on the firm's performance. Moreover, we note that not only portfolio of hedges are more effective than single hedges, but the corresponding optimal decisions are easy to compute as long as the futures contract is included in the portfolio. The inventory decisions are myopic and the hedging quantities can be easily determined by solving a system of linear equations. The above observations back up our recommendation for the use of portfolio of hedges in practice.

Lastly, we discuss the observations from the numerical study on the impact of the correlation between demands and spot prices (reflected by $\rho = -0.5, 0, 0.5$). As Figures 6.3 and 6.4 demonstrate, the operational hedge (the spot market procurement) is most effective at $\rho = -0.5$ and least effective at $\rho = 0.5$; the financial hedge, however, is most effective at $\rho = 0.5$

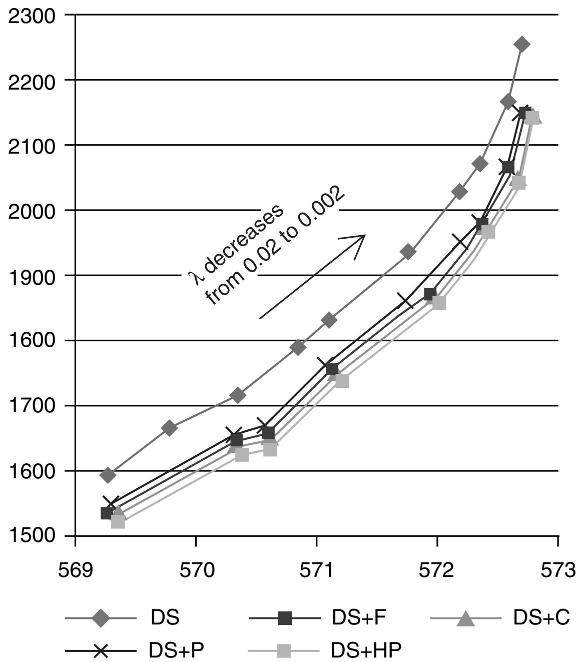


FIGURE 6.5 Efficient frontier (variance versus mean) at $\rho = -0.5$.

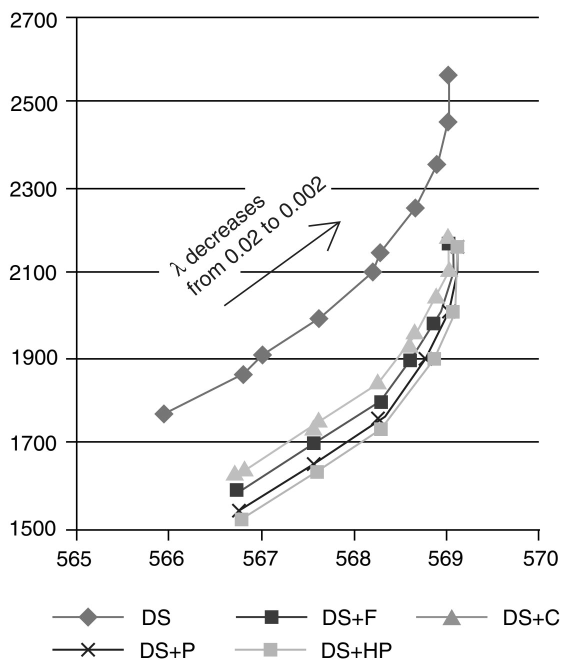


FIGURE 6.6 Efficient frontier (variance versus mean) at $\rho = 0.5$.

and least effective at $\rho = 0$. The result for the financial hedge is not surprising. For the operational hedge, we note that a negative correlation means higher demands occur with lower spot prices, which translate to a higher profit margin. This may explain why the operational hedge is most effective in the negative correlation case.

6.7 Example of Model Application and Results

We give two numerical examples, one simple with discrete demand and commodity price and one practical with continuous demand and spot price, to illustrate how the optimal inventory and hedging decisions are determined and their relationship.

Let us first examine a simple two-period model with zero supply from the long-term contract (i.e., $q = 0$) and discrete price and demand. The sequence of events, including decision and observed information, is shown in Figure 6.7, where S_i represents commodity price at time i, $i = 0, 1, 2$, and D_i represents demand in period i, $i = 1, 2$.

The commodity price dynamics and the demand dynamics are illustrated by Figures 6.8 and 6.9, respectively.

Our model assumes that the firm's objective at each decision time epoch is to maximize his mean-variance utility of the to-go cash flow, where the cash inflow consists of sales revenue and payoff of the financial hedge and the cash outflow includes the procurement cost and inventory holding cost. For this simple example, we assume that there is no cash flow discount over time (i.e., $\alpha = 1$), unit sales revenue $r = \$15$, unit holding cost $h = \$3$, and the risk aversion factor $\lambda = 0.001$. For this example, we observe that for the second period, it is optimal to use the futures contract alone. That is myopic hedging is optimal. If the commodity price S_1 is high (e.g., \$12), the firm should keep the inventory Z_2^* at the minimum level, that is the minimum demand, 80 units. If, on the other hand, the commodity price is low (e.g., \$8 and \$10), the firm should keep the inventory at the maximum level, that is the maximum demand, 120 units. The details are shown in Table 6.1.

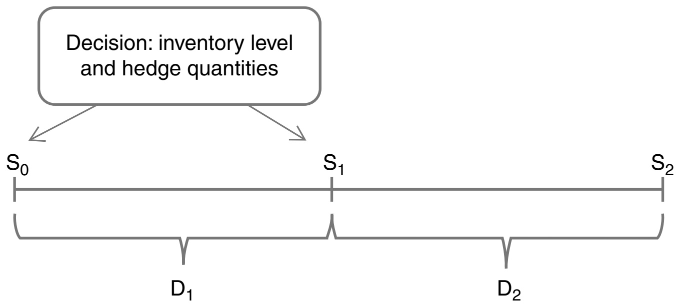


FIGURE 6.7 Sequence of events illustration for a two-period model.

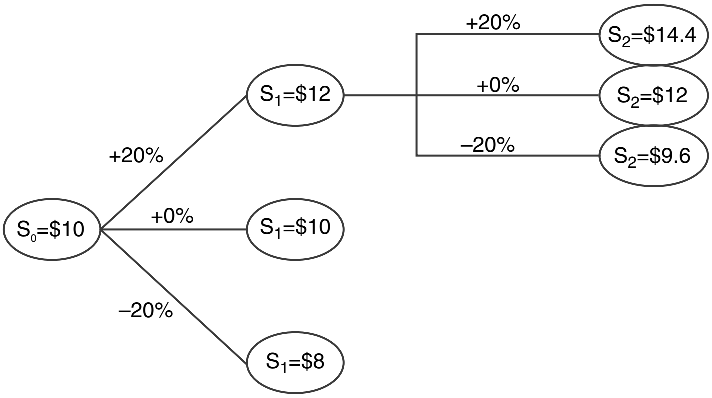


FIGURE 6.8 Commodity price dynamics (with equal probabilities).

We observe similar behavior for inventory decisions for the first period. Since the commodity price s_0 is low, the firm should keep the inventory at the maximum level, 120 units. For this period, however, myopic hedging is not optimal. For illustration purposes, we compare two hedges. Hedge I, the myopic hedging, contains the futures contract only, while Hedge II contains the futures contract and a call option with strike price \$9 and maturity time start of period 2. Note that, under fair pricing for the financial contracts, the choice of hedging does not change the mean cash flow. In other words, the choice of hedging only affects the variance of the cash flow. We find that the variance of the total cash flow in two periods is reduced by 96.2% when hedge II is used, compared to that when

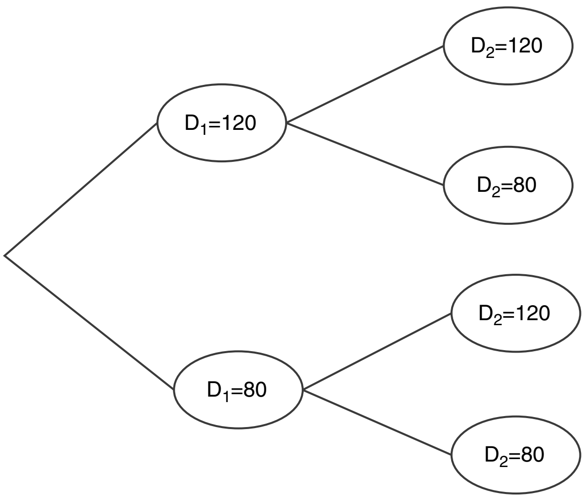


FIGURE 6.9 Demand dynamics (independent with equal probabilities).

TABLE 6.1 Optimal Inventory and Hedging Decisions for Each Period

	S_1	Z_2^*	Optimal Hedge: $y_{f,2}^*$	
Period 2	\$12	80	0	
	\$10	120	20	
	\$8	120	20	
	S_0	Z_1^*	**Hedge I (myopic):** $y_{f,1}^*$	**Hedge II:** $y_{f,1}^*, y_{c,1}^*$
Period 1	\$10	120	80	800, −960
	Variance of total cash flow		230,627.6	8,760.9 (down by 96.2%)

hedge I is used. This example clearly indicates the importance of choosing the right hedging portfolio, as well as that myopic hedging (i.e., using the futures contract alone in our model) may be far from optimal.

We next state the setup of a two-period numerical study with identical periods (and thus we remove the subscript n for the cost parameters) and correlated spot prices and end-product demands, both follow Lognormal distribution or Geometric Brownian motion (GMB) when treated as continuous time stochastic processes. The results and figures from this study have been presented in previous sections of the chapter. Specifically, let $S_{n+1} = \frac{S_n}{\gamma} e^{\sigma_s B_{n+1} - \sigma_s^2/2}$ and $D_n = \mu e^{\sigma_d W_n - \sigma_d^2/2}$, where B_n's and W_n's are iid standard Normal random variables, except that (B_n, W_n) is bivariate Normal with correlation $\rho \in [-1, 1]$, and $n = 1, 2$. In order to choose reasonable values for parameters for the spot price, γ and σ_s, we fit the spot price distribution to the U.S. hot-rolled coil steel spot price from January 2009 to June 2010. The fitting indicates that the steel spot price is approximately GMB with monthly discount factor $\gamma = 0.9758$ and monthly drift parameter $\sigma_s = 0.114$. Since we are interested in "fairly" comparing various financial hedges, we assume zero risk premium for the numerical study and reflect this assumption by letting $\alpha = \gamma = 0.9758$ (which implies the buyer's expected rate of return equals to the risk-free rate of return, i.e., $r = r_f$). In particular, we perform the computational study with the following parameters: $\alpha = 0.9758$, $w = 3$, $q = 60, 80, \ldots, 200$, $s_1 = 3$, $r = 4, 5, 6, 7, 8$, $h = 0.6$, $\sigma_s = \sigma_d = 0.114$, $\rho = -0.5, 0, 0.5$, $\mu = 100$, $\lambda = 0.002, 0.006, 0.01, 0.02, 0.06, 0.1$.

Note that the range of λ is set appropriately small enough to avoid a pathological behavior of MV decision makers—reducing inventory when seeing higher revenue as it drives up sales and thus profit variance. Our analysis show that z_n^* increases as r increases if $\lambda \le \frac{1}{2(r+h-s_n)E\left[\left(\bar{z}_n^* - D_n\right)^+\right]}$, where $\bar{z}_n^* = F_n^{-1}\left(\frac{r-s_n}{r+h-s_n}\right)$ represents the risk-neutral optimal inventory level for period n. Applying this to our numerical study setup with say $r = 6$, we obtain that the buyer stocks more as the unit revenue rises if his risk aversion is below 0.114.

6.8 Managerial Insights and Conclusions

When sourcing input commodities with volatile prices and processing them to meet volatile end-product demand, the manufacturers are searching for effective ways to manage the resulted volatile cash flows. The concerned industries are as diverse as food processing, autos, household products, and hi-tech electronics. Recent research reported in Kleindorfer (2008 and 2010) has emphasized the need for integrating long-term bilateral contracts (fixed commitment or flexible) with access to reasonably liquid spot markets. Furthermore, as financial intermediaries nowadays offer a variety of financial contracts (futures, call and put options) written on commodity spot prices, there are many means to effectively hedge the risk exposure in commodity procurement decisions. Our paper offers an integrated risk management framework for storable commodities deploying dual sourcing via a fixed price, fixed quantity long-term supply contract and short-term commodity exchange purchases/sales combined with a portfolio of financial hedges for cash flow volatility control. Our results offer useful insights on how much to source from the spot market, optimal inventory policies in the presence of both end-product demand and commodity price uncertainties, which could be correlated, and structure of the optimal financial hedging portfolio.

Effective implementation of such policies requires cross-functional decision coordination and information sharing among operations and financial managers. Our results indicate that the information burden is lower for operations managers in effectively executing sourcing and inventory decisions. The setting of optimal base stock levels requires awareness of the firm's commitment to financial hedging and the type of hedging contract to be used, without however requiring the details of the hedging quantities. The optimal base stock level is myopic and the same for all portfolios of hedges that include futures, and the base stock calculations are given in closed-form formulas (see Proposition 6.2 on inventory). For such cases, the base stock levels are decreasing in risk aversion, observed spot market prices, and the variance of spot market prices.

The financial manager implementing an integrated commodity risk management approach is required to know the base stock levels for effectively choosing the hedge type and amounts for the optimal hedging portfolio. The good news is that the optimal financial hedging portfolio is obtained via the straightforward solution of a system of linear equations (see Proposition 6.2 on hedging). For short-term horizons, more or less resembling our single (last) period results, futures contracts are all that are needed to hedge the cash flow variance. However, this is not the case for longer horizons (unless we are working with deterministic demand forecasts, rather than random demands, for future periods). The institutional reality of many commodity markets (lack of financial intermediaries offering option contracts with enough liquidity) makes the use of futures contracts as the only practical solution. For that case the optimal inventory levels and hedge amounts are described in closed forms in Propositions 6.1 and 6.2.

Our work clearly shows the role played by the operational and financial hedges. Our two operational hedges, the long-term contract and access to the spot market, are effectively used to deal with the demand uncertainty. By its nature, the long-term contract protects the buyer against spot market price volatility, while the financial derivative contracts hedge directly the spot price uncertainty and only indirectly the demand uncertainty. The operational hedge is similar to, but more effective than, the financial hedge if only one is allowed for use. When jointly used, they are complementary to each other, each with distinct roles, where the operational hedge focuses on improving the mean cash flow and the financial hedge controls the cash flow volatility. For single period settings, the use of financial hedges increases inventory and improves service level. Finally, the inventory levels decrease as the buyer becomes more risk averse. The observations are still true for zero risk premium environments even for multiperiod settings. However, these observations will not hold for nonzero risk premium and multiperiod problems.

REFERENCES

Alexander J.C., Sobel M.J. 2006. Risk neutrality and ordered vector spaces. Cleveland, Ohio: Case Western Reserve Technical Report No. TM-808A.

Basak S., Chabakauri G. 2010. Dynamic mean-variance asset allocation. *The Review of Financial Studies*, **23**(8):2970–3016.

Bodily S., Palacios M. 2007. Integrating financial hedges and operational options in an electric power plant. Working paper. Berkeley, California: UC Berkeley.

Boyabatli O., Kleindorfer P.R., Koontz S. 2011. Integrating long-term and short-term contracting in beef supply chains. *Management Science*, Forthcoming.

Bryan L. 2010. Globalization's critical imbalances. *McKinsey Quarterly*, June 2010.

Blas J., Gorst I. 2010. Rise in wheat prices fastest since 1973. *Financial Times*, August 2, 2010.

Cachon G. 2003. Supply chain coordination with contracts. In: Graves S., de Kok T., editors. *Handbooks in Operations Research and Management Science: Supply Chain Management*. Amsterdam: Elsevier/North Holland.

Carr P., Madan D, 2001. Optimal positioning in derivative securities. *Quantitative Finance*, **1**:19–37.

Chod J., Rudi N., Van Mieghem J.A. 2010. Operational flexibility and financial hedging: Complementarity or substitutes? *Management Science*, **56**(6):1030–1045.

Ding Q., Dong L., Kouvelis P. 2007. On the integration of production and financial hedging decisions in global markets. *Operations Research*, **55**(3):470–489.

Dothan M.U. 1990. *Prices in Financial Markets*. New York: Oxford University Press.

Fisher B., Kumar A. 2010. The right way to hedge. *McKinsey on Finance*, 36.

Froot K.A., Scharfstein D.S., Stein J.C. 1993. Risk management: Coordinating corporate investment and financing policies. *The Journal of Finance*, **48**(5): 1629–1658.

Geman H. 2005. *Commodities and Commodity Derivatives*. New York: Wiley.

Goel A., Gutierrez G.J. 2006. Integrating commodity markets in the optimal procurement policies of a stochastic inventory system. Working paper. Cleveland, Chio: Case Western Reserve University.

Jucker J.V., Carlson R.C. 1976. The simple plant-location problem under uncertainty. *Operations Research*, **24**:1045–1055.

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

Kleindorfer P.R. 2008. Integrating physical and financial risk management in supply management. In: Geman H., editor. *Risk Management for Commodity Markets*. New York: Wiley.

Kleindorfer P.R. 2010. Hedging commodity risks in supply management. Note in: Rama C., editor. *Encyclopedia of Quantitative Finance*. New York: Wiley.

Kouvelis P., Li R., Ding Q. 2010. Managing storable commodity risks: Role of inventories and financial Hedges. Working paper. Singapore Management University.

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

Nagali V., et al. 2008. Procurement risk management (PRM) at Hewlett-Packard Company. *Interface*, 38, 51–60.

Oum Y., Oren S., Deng S. 2006. Hedging quantity risk with standard power options in a competitive wholesale electricity market. *Naval Research Logistics*, **53**:697–712.

Peleg B., Lee H.L., Hausman W.H. 2002. Short-term E-procurement strategies versus long-term contracts. *Production and Operations Management*, **11**(4):458–479.

Smith J.E., Nau R.F. 1995. Valuing risky projects: Option pricing theory and decision analysis. *Management Science*, **41**(5):795–816.

Tevelson R.M., Ross J., Paranikas P. 2007. A key to smart outsourcing: Hedging against risk. Boston, Massachusetts: Boston Consulting Group Report.

Wiggins J. 2008. Hedging is second nature for food groups. *Financial Times*, August 17, 2008.

Wiggins J., Blas J. 2008. Food and consumer groups get hedge smart. *Financial Times*, August 17, 2008.

Wu D.J., Kleindorfer P.R. 2005. Competitive options, supply contracting, and electronic Markets. *Management Science*. **51**(3):452–466.

Wu D.J., Kleindorfer P.R. Zhang J.E. 2002. Optimal bidding and contracting strategies for capital-intensive goods. *European Journal of Operational Research*, **137**(3):657–676.

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

Van Mieghem J.A. 2007. Risk mitigation in newsvendor networks: Resource diversification, flexibility, sharing, and hedging. *Management Science*, **53**(8):1269–1288.

CHAPTER SEVEN

Integrated Production and Risk Hedging with Financial Instruments

ÇAĞRI HAKSÖZ AND SRIDHAR SESHADRI

We review the existing literature on integrated production and risk hedging with forwards/futures and options for a risk averse firm in single and multiperiod settings. We illustrate the value of hedging joint price, basis, and yield risks using forwards/futures and options. We then focus on a procurement problem for a risk neutral commodity producer who sells to its buyer (with a stochastic demand) via a long-term fixed-price contract, and trades intelligently in the spot market for the commodity. We solve a continuous time, infinite horizon stochastic control problem in order to determine the optimal policy for production and spot market trading.

Acknowledgements: We are grateful for the suggestions of the editors and the referees which enhanced the quality and the exposition of the chapter. We also kindly acknowledge the beneficial discussions with Vishal Gaur, Victor Araman, René A. Caldentey, Qi Wu, and Akın Sayrak.

Handbook of Integrated Risk Management in Global Supply Chains, First Edition. Edited by Panos Kouvelis, Lingxiu Dong, Onur Boyabatli, and Rong Li.

7.1 Introduction

Today's highly uncertain and interconnected world requires manufacturers all over the globe to smartly manage their supply chain risks (Tang 2006 and Sheffi 2007). More markets open and become resource-hungry, and supply networks enlarge and span many more countries and continents. Supply disruptions (Dada et al. 2007, Tomlin 2009), supplier bankruptcies/defaults (Babich et al. 2007), supply contracts breaches (Haksöz and Kadam 2008 and 2009), product recalls (Tang 2008, Sezer and Haksöz 2010), and network breakdowns become ubiquitous. In such a world, manufacturers have been contemplating better and more effective ways to hedge their interdependent risks such as demand, market price, and production yield.

Risk hedging is mainly used to minimize the variance of firms' cash flows (Van-Mieghem 2007). Financial instruments, if used correctly and carefully, provide useful avenues for effective risk hedging, by reducing the downside exposures of global manufacturers (Gaur and Seshadri 2005, Caldentey and Haugh 2006, Chen et al. 2007, Ding et al. 2007, Tevelson et al. 2007, Chod et al. 2010). Price risk can be hedged by using forward and futures contracts, and trading in the options markets. On the other hand, spot market trading can be used for hedging demand risk. Especially, erratic price history of spot markets expose commodity producers and industrial users such as automotive, high-tech, chemicals, fast moving consumer goods, and pharmaceutical firms to much higher risks than in the past. We need to mention that the spot price is driven by the supply (mostly from the producers) and the demand (mostly from the buyers). Moreover, the spot price may also be affected by the use of different subjective probability distributions and subjective spot price forecasts by different producers and buyers.

Matthews (2008a) reports that ArcelorMittal, the largest steelmaker in the world, sold only 20% of its steel output via contracts, and the rest was sold directly on the spot market to take advantage of high prices. However, by the end of 2008, as the global economic crisis worsened, commodity prices plummeted around the world. With the decline of crude oil price from \$147 a barrel in mid-July 2008 to below \$50 in December 2008, several other commodities prices such as steel, wheat, and ethylene fell by 68%, 67%, and 50% respectively during the March 2008 to August 2008 period as reported in the World Economic Forum *Global Risks 2009: A Global Risk Network Report* (2009). These price drops in the market turned the buyers' interest to the spot markets in order to avoid paying much higher negotiated contract prices to the producers. For example, Matthews (2008b) presents the case of the Australian iron-ore producer Mount Gibson Iron Ltd. in which three of Mount Gibson's customers defaulted on their contracts to purchase from Mount Gibson and resorted to purchasing iron ore in the spot market. Hence, spot market trading can be strategically used in both high price as well as low price regimes depending on who is trading in the market (producer or buyer).

The contents of the chapter will include the following: We first review the existing literature on integrated production with risk hedging using financial instruments such as forwards/futures, and options. Beginning with single-period models and one type of risk (i.e., market price risk), we show the value of hedging with forwards/futures as well as options. Then, we demonstrate the implications of basis and yield risks on optimal production and hedging decisions. Basis risk occurs when there is mismatch between the futures and the spot price at the expiration date of the futures contract. Yield risk refers to the production uncertainty. Later, we examine the models to date in order to demonstrate the role of hedging with futures/forwards and options in multiperiod settings. Multiple uncertainties such as basis, yield, and demand risks will be considered. In addition to the review, we study a specific problem that has many applications in practice. In this problem, a commodity producer sells to fully satisfy the stochastic demand of the buyer via a fixed-price long-term contract while minimizing the total costs. We incorporate spot market trading into the production planning and determine its value for the producer under various production and spot market conditions.

A previous review article by Haksöz and Seshadri (2007) demonstrated that understanding the value of spot market trading under different procurement models would be beneficial. In Haksöz and Seshadri (2007), a fixed-price long-term contract with an abandonment option given to the producer is valued in a complete and frictionless market setting. In this chapter, we study the case where contract abandonment may not be possible due to various strategic, operational, and legal reasons such as the possibility of reputation/credibility risks in case of abandonment, or the likelihood of future deal losses with the buyer. In some cases, the producer may be exposed to time-consuming and costly litigation. When the contract excludes the abandonment option, the producer becomes more interested in using the spot market to buy and sell commodities. The buyer's demand may also fluctuate within the contract horizon, so optional use of spot market becomes more critical in case of higher/lower demand realizations. Thus, although the producer has promised to sell a certain production capacity within a range via the long-term contract, she may seek to earn extra profits by spot market trading. This is the problem we study in Section 7.3.3 of this chapter. In essence, we allow the producer to optimally control production and the spot market trading in the presence of potentially correlated *spot price* and *demand* risks. To that end, we also demonstrate the value of trading in the spot market for the commodity producer.

The chapter is organized as follows: Section 7.2 presents the single period production and risk hedging models with price, basis, and yield risks. Section 7.3 presents models in multiperiod settings. In Section 7.3.3, we present a continuous time stochastic control model in order to determine the optimal production quantity and spot market trading for a risk neutral producer. Key managerial levers are identified and their impact on spot market trading are presented. We conclude with managerial implications in Section 7.4.

7.2 Single Period Models

In this section, we review single period models for a risk averse firm. The firm decides on the production volume as well as how much price risk to hedge using financial instruments. Later, we address the joint price, basis, and yield uncertainties. We first demonstrate how the firm behaves in the absence of any hedging capability. Then, we demonstrate the value of hedging using forwards/futures and options with illustrative examples.

7.2.1 PRODUCTION PLANNING UNDER PRICE UNCERTAINTY

First, we present the existing models that incorporate a single uncertainty (i.e., spot or market price). Later, we will present models that consider the impacts of basis and yield risks.

In this part, we begin with a simple model that computes the optimal production levels in the presence of price risk. Hedging with financial derivatives is not considered. The firm is risk averse and maximizes the expected utility. After summarizing the main results we illustrate the computation and details using a simple example.

In this line of research, Sandmo (1971) studies the theory of a competitive firm under price uncertainty and risk aversion. The firm is a price taker and the demand is not known with certainty when the production decision is taken. The price uncertainty is modeled via the subjective distribution of the prices. The firm maximizes the short run expected utility of profits to determine the optimal level of production prior to the date when price would be realized. As is well known now, Sandmo (1971) shows that the production level is lower under price uncertainty than the case when the price is certain. Sandmo (1971) also examines the effect of increase in price risk on the production level. We do not discuss these results because they are subsumed by the ones due to Batra and Ullah (1974) reported below. The long-term impact of risk aversion in equilibrium is also studied by Sandmo (1971). He shows that risk neutral firms enter the industry as long as their profits are non-negative, hence the marginal cost is equal to the expected price. Yet, risk averse firms will choose to produce at the levels where the marginal cost is less than the expected price, thus the most risk averse firms will be reluctant to enter the market. Sandmo (1971) also reports results of comparative statics to demonstrate the impact of fixed costs and tax rates on the production level. It is important to note that effects are obtained under risk aversion and would not be obtained under risk neutrality.

Batra and Ullah (1974) generalizes the seminal work by Sandmo (1971) by incorporating two factors of production (capital and labor) in a long run decision making setting. The firm being a price taker maximizes the expected utility of profits by deciding on the optimal production level before the price is known with certainty. Thus, the firm decides upon the level of capital and labor prior to observing the price. The main results of the paper are as follows: For risk averse firms, production function concavity is sufficient but not necessary for the firm to reach the optimal solution. For risk neutral firms, however, production function concavity is necessary and sufficient. For risk seeking firms, this condition is necessary, but not sufficient. In the presence of price uncertainty, a risk averse firm decreases the production level. Moreover, as the price uncertainty increases via a mean preserving spread (see Rothschild and Stiglitz (1970) for a detailed discussion), the firm reduces the production level if the firm has decreasing absolute risk aversion.[1] Batra and Ullah (1974) show that as the expected price increases, risk neutrality, decreasing absolute risk aversion or increasing absolute risk aversion are sufficient to induce the firm to increase its production output. An increase in input costs induces the firm to reduce the demand for that input (capital, labor) and thus also reduces the production level. Yet, it is not predictable how the demand for other input will vary while changes occur in the other input.

Example 7.1

Consider that the firm has constant absolute risk aversion (CARA) utility function which is given as $U(\Pi) = -e^{-\alpha\Pi}$, $\alpha > 0$. Measure of risk aversion is $r(\Pi) = \alpha$.[2] The firm maximizes the expected utility with strictly convex production costs, $c(x)$. The profit is computed in (7.1) where $\tilde{p}$ is the random market price at which the firm sells its production output x. Please note that $\tilde{}$ and $\bar{}$ denote random variables and mean of random variables respectively throughout the chapter. Assume the price $\tilde{p}$ follows Normal distribution with mean μ and standard deviation σ. Let $f()$ denote its probability density function (pdf). The firm wants to maximize the expected utility by choosing an optimal production quantity x.

$$\tilde{\Pi}(x) = \tilde{p}x - c(x) \tag{7.1}$$

$$\underset{x\geq 0}{\text{Max}} EU(\tilde{\Pi}) = \underset{x\geq 0}{\text{Max}} \int_{-\infty}^{\infty} U(\tilde{p}x - c(x)) f(\tilde{p}) d\tilde{p} \tag{7.2}$$

The optimization problem becomes maximization of

$$\mu x - c(x) - \frac{\alpha x^2 \sigma^2}{2} \tag{7.3}$$

Thus, the first order condition (FOC) yields the optimal amount of production, x^*, where * denotes the optimal values of the variables, in the absence of hedging such that the (7.4) holds:

$$c'(x^*) = \mu - \alpha x^* \sigma^2 \tag{7.4}$$

---

[1] Gollier (2001) states that if a firm has decreasing absolute risk aversion, then, the more wealthy the firm is, the smaller is the maximum amount it will pay to avoid a given additive risk.

[2] We should note that CARA is not the only risk aversion characterization of real-life agents. Decreasing absolute risk aversion (DARA) and increasing absolute risk aversion (IARA) are also used in the literature for exponential utility functions. Basically, in a general exponential utility function, the absolute value of the coefficient of Π in the power is defined as the absolute risk aversion, which measures the risk and can be constant, increasing or decreasing with respect to the profit Π. Surely, not all attitudes of agents can be modeled with these risk aversion assumptions. Nevertheless, they are used for ease of analytical manipulation. For details on risk aversion, see for instance Chapter 2 in Gollier (2001).

Note that $c'(0) < \mu$, otherwise the optimal production quantity $x^* = 0$.

As the risk aversion coefficient, α increases, the firm decreases the optimal production output. This causes the firm to produce such that $c'(x^*) < \mu$. Note that in the comparable case with a deterministic market price (i.e., $\tilde{p} = \mu$), the FOC is $c'(x^*) = \mu$ and the profit is $\mu x^* - c(x^*)$. Therefore, the comparison implies that $\frac{\alpha x^2 \sigma^2}{2}$ can be considered as the risk premium, paid to the Normal market price. Thus, the firm reduces the production output from the level that maximizes the expected profit to reduce the variance.

We could also examine the impact of an increase in riskiness of the price distribution on the optimal production output. Therefore, let us compute the second order condition (SOC) of the expected utility as follows:

$$D = E[U''(\tilde{\Pi})(\tilde{p} - c'(x))^2 - U'(\tilde{\Pi})c''(x)] < 0 \tag{7.5}$$

This ensures that the FOC yields the global maximum.

Let us assume that the increasing risk in the price distribution is achieved via a mean preserving spread. Mean preserving spread is defined by Sandmo (1971) as stretching the probability distribution around a constant mean. In other words, the probability distribution of price is made slightly more risky. We have two shift parameters, one is multiplicative shift, γ, the other is the additive shift given by θ. Thus, we can write the price as:

$$\gamma \tilde{p} + \theta \tag{7.6}$$

To prevent the blow up of the price (by increasing the γ alone), that is, it will increase both the mean and the variance, we need to have hold:

$$\mu d\gamma + d\theta = 0 \tag{7.7}$$

Note that this implies $d\theta / d\gamma = -\mu$.

In order to decipher the impact of price risk on production, we differentiate the profit function $\Pi(x) = (\gamma \tilde{p} + \theta)x - c(x)$ with respect to γ, considering $d\theta / d\gamma = -\mu$, and we obtain

$$\frac{\partial x}{\partial \gamma} = -x\frac{1}{D}E[U''(\tilde{\Pi})(\tilde{p} - \mu)(\tilde{p} - c'(x))] - \frac{1}{D}E[U'(\tilde{\Pi})(\tilde{p} - \mu)]. \tag{7.8}$$

If $c'(x^*) = \mu$ holds, then the first and the second term in the equation above becomes negative. The overall sign of the derivative depends only on the risk aversion assumption.

One can also introduce two inputs (capital and labor) into this model following the analysis by Batra and Ullah (1974). Assume that the production output is a function of capital and labor used. Hence, where K is the capital and L is the labor'.

$$x = g(K, L) \tag{7.9}$$

Also assume that the firm has a certain knowledge of the capital and labor price, r and w. $c(x) = wL + rK$ is a linear function of the capital and labor costs. The firm's profit function can be now written as

$$\tilde{\Pi}(x) = \tilde{p}g(K, L) - wL - rK \tag{7.10}$$

Computing the FOC and SOCs, one can obtain the conditions under which the concavity of production function becomes necessary and sufficient for expected utility maximization.

To show the impact of price uncertainty on the optimal production output and the input quantities, we use the FOCs, which are given by (7.11) and (7.12), where g_L and g_K are partial derivatives of labor and capital respectively and $U'(\tilde{\Pi}) = \frac{dU(\tilde{\Pi})}{d\tilde{\Pi}}$.:

$$E[U'(\tilde{\Pi})\tilde{p}g_L] = E[U'(\tilde{\Pi})w], \tag{7.11}$$

and

$$E[U'(\tilde{\Pi})\tilde{p}g_K] = E[U'(\tilde{\Pi})r] \tag{7.12}$$

If the utility function is strictly concave, $U'' < 0$, subtracting $E[U'(\tilde{\Pi})\mu g_L]$ from both sides of the first FOC and subtracting $E[U'(\tilde{\Pi})\mu g_K]$ from both sides of the second FOC and manipulating algebraically, one can show that $w \leq \mu g_L$ and $r \leq \mu g_K$ hold. This means that with price uncertainty, the optimal input quantities will be lower than those in the certainty case. Hence, the optimal production output will be also lower.

7.2.1.1 Hedging with Forwards/Futures. Johnson (1960) sets up a theoretical framework to explain the determinants of futures contracts hedging and spot market trading. First, he defines the "hedger" to be a dealer in the actual commodity who requests insurance against price risk. Futures markets are used for transferring this price risk from one group to the other, mainly from hedgers to speculators. Thus, hedgers are willing to pay a risk premium to obtain risk mitigation, whereas speculators aim to collect this risk premium while trading in the futures market. Johnson (1960) also mentions the work by Working (1953) who sees hedging not only as a risk avoidance function, but also a function to obtain returns due to favorable price movements in spot and futures markets. Moreover, hedging effectiveness is based on the inequalities between the spot and futures price movements and the predictability of these inequalities.

Johnson (1960) defines the hedge as taking positions in a specific market for a certain duration given a specific position in another market such that price risk is minimized. For example, having open (or uncovered) position in the spot market, the decision maker can have long or short futures positions to hedge the price risk exposed due to the spot market position. Later, Johnson (1960) analyzes hedging and speculation in two markets (spot and future markets) in a simple, yet effective model. In the model, there are no budget constraints, interest rates or brokerage commissions. Expected return of the total return is plotted against the variance of the total return. It is shown that the hedge effectiveness depends on the correlation coefficient between the spot and futures price changes. In other words, the trader believes that the effectiveness is measured by the extent to which the variance of return holding a position in one market is reduced by holding another position in the other market simultaneously. Thus, primary market has to be specified in any hedging computation in which the 'other' market position needs to be determined.

In his geometrical analysis, Johnson (1960) demonstrates the following: Opportunity lines can be drawn for different decision makers that show the combination of spot market positions and the futures positions as a hedge. In this setup, a pure hedging, direct speculation, and indirect speculation cases can be demonstrated. Pure hedging occurs when a trader takes a futures hedge position to minimize the price risk due to the position taken in the spot market. Direct speculation occurs when a position is taken in the primary market for an expected nonzero price change in the same market. Indirect speculation occurs when an expected nonzero price change exists in the other market. For instance, an expected increase in the spot market price may induce the trader to increase his position by purchases in the spot market, which is a direct speculative element. In addition, this increase will induce the trader to change his position in the short futures providing a hedge against this increase in the spot market, which is an indirect speculative element. On the other hand, Stein (1961) constructs the optimal allocation of hedged and unhedged stocks to maximize the expected utility. Hedging is done via futures contracts. In the process, he describes two effects. One is the *substitution effect*, where the firm increases the ratio of hedged stocks to total stocks as the price of futures contracts increases. The offsetting effect is the *income effect*; as the firm obtains higher expected utility, it tends to be less risk averse and thus reduces the hedging component. The spread, which is defined as the difference between the spot and the futures contract price today, is increased by the excess demand for futures contracts.

Holthausen (1979) and Feder et al. (1980) study a single period model of a risk averse firm that can engage in forwards/futures trading in addition to production in the presence of price uncertainty. In this model, the risk averse firm decides on the optimal production output as well as the futures trading volume simultaneously. The firm maximizes the expected concave utility function. At the time of decision, the futures contract price is known, yet the terminal futures spot price is not known with certainty. Only its distribution is known. A general concave production function is used to determine the output.

As Sandmo (1971) and Batra and Ullah (1974) showed, an increase in price uncertainty (given by mean preserving spread of the price distribution), induces the firm to reduce production. Yet, increase in the expected price raises the production level. In contrast, in the presence of forwards/futures trading, production decision is affected neither by the subjective distribution of the future spot price, nor by the firm's risk aversion. Production is positively affected by the prevailing futures price and negatively affected by the production cost. This is the famous separation property between production and the futures trading. Also note that this result holds for commodities whose production uncertainty is nonexistent. Further, if the expected price is equal to the forward price, the firm will hedge its entire production. If the expected price is higher than the forward price, the firm will either hedge less than its entire production or if the forward price is sufficiently less than the expected price, the firm will speculate by purchasing production in the forward market, expecting to sell it later at a higher price. If the expected price is lower than the forward price, the firm will speculate by selling more than its production, expecting to purchase at lower prices in the future and pocket the difference.

If the forward price is less than the expected price, as the risk aversion increases, the firm will hedge a higher amount. On the other hand, if the forward price is higher than the expected price, the more risk averse the firm becomes, the less it speculates. The amount of forward hedging increases as the riskiness of the probability distribution of price increases (with mean preserving spread) for a non-increasing absolute risk averse firm. Presence of forward or futures markets induces the firm to increase production level, *ceteris paribus*, given the optimal hedge is positive. If the optimal hedge is negative, production will be reduced in the presence of hedging.

From a different perspective, Lence (1995) addresses the fundamental question whether minimum variance hedges (MVH) continue to hold under realistic assumptions and moreover whether there is significant economic value in better MVH estimates. In his paper, a decision maker having concave utility function with CARA maximizes the expected value of terminal wealth in a single period setting. The decision maker optimally chooses four actions which are the amount of commodity to sell in the futures market, amount of money lent and borrowed at different rates of return, and amount of money invested in a risky instrument. The model handles realistic situations, which are hitherto overlooked in previous work, safety margins deposited for futures trading, futures trading fees, and a budget constraint determining how much the decision maker can borrow.

The optimal hedge in this model is found using standard Lagrangian multiplier technique. In general, if the objective function has certain restrictions such that no lending/borrowing/investment in alternative activities allowed as well as no safety margins nor futures trading fees, MVH decisions are similar to those of expected utility maximizers. Yet, in the presence of no restrictions, MVHs become suboptimal. Lence (1995) measures the value of this suboptimality by using opportunity costs. Opportunity cost is defined as the maximum certain net return the decision maker is willing to pay for the right to invest in the optimum strategy rather than in the alternative investment. In other words, it is the minimum certain net return required by the decision maker to invest in the alternative investment rather than the optimum investment. Furthermore, Lence (1995) also studies the estimation risk, which is the uncertainty regarding the probability distribution of the returns. Using Bayesian theory, one can use prior distributions and update the distribution parameters as more information is obtained in time.

Lence (1995) solves the model numerically using further assumptions such as: CARA utility function, random returns for the commodity sold at terminal date in cash and futures markets and the random return of the alternative investment have multivariate normal distributions. He compares the optimal expected utility versus the expected utility using MVHs for hedging. The simulation results demonstrate the following: The optimal hedge ratio depends sensitively on the common restrictions. Under realistic assumptions, where the decision maker is allowed to lend, borrow, and invest in alternative instruments rather than cash and futures positions, the optimal hedge ratios become zero. Increasing costs of trading with futures reduce the optimal hedge ratios. Opportunity costs of small deviations from the optimal hedge ratios are almost negligible. Thus, potential benefits by reducing such deviations are limited. The value of having better MVH estimates is minuscule. Thus, he concludes that the opportunity cost of learning the true estimate of correlation between the cash and the futures markets is small, while finding the optimal hedge ratios. These results suggest that it might be alright to use MVH when one really has CARA utility function while solving the joint production and hedging problem.

Example 7.2

Given the same assumptions for the utility function, production costs and the price distribution presented in the Example 7.1, the firm now hedges the price risk using forward contracts. Following the results by Holthausen (1979), we assume that the firm has the option to trade in the forward market at given certain price b. The quantity of forward hedging is given by h. The profit function will be then given by

$$\tilde{\Pi}(x, h) = \tilde{p}(x - h) + bh - c(x). \tag{7.13}$$

The firm maximizes the expected utility function while deciding on the optimal production level x and the forward hedging quantity h. Thus we have

$$\underset{x,h\geq 0}{\text{Max}} EU(\tilde{\Pi}) = \underset{x,h\geq 0}{\text{Max}} \int_{-\infty}^{\infty} U(\tilde{p}(x - h) + bh - c(x)) f(\tilde{p}) d\tilde{p} \tag{7.14}$$

The FOCs are given by

$$\frac{\partial EU(\tilde{\Pi})}{\partial x} = \int_{-\infty}^{\infty} U'(\tilde{\Pi})(\tilde{p} - c'(x)) f(\tilde{p}) d\tilde{p} = 0 \tag{7.15}$$

and

$$\frac{\partial EU(\tilde{\Pi})}{\partial h} = \int_{-\infty}^{\infty} U'(\tilde{\Pi})(b - \tilde{p}) f(\tilde{p}) d\tilde{p} = 0 \tag{7.16}$$

Assuming CARA and normality for the price distribution, we obtain the following expression that is to be maximized:

$$\mu(x-h) + bh - c(x) - \alpha(x-h)^2\sigma^2/2 \tag{7.17}$$

If one compares expression (7.17) with (7.3) given in Example 7.1, (7.17) has the additional term due to the forward trade, bh. Besides, the production level, x, and the risk premium, $\alpha x^2\sigma^2/2$ are reduced with the addition of forward hedging quantity h, thus becoming $(x-h)$ and $\alpha(x-h)^2\sigma^2/2$ respectively. The FOCs yield the following two expressions such that the firm would produce and hedge via forwards:

$$c'(x^*) = b, \;\; x^* - h^* = \frac{\mu - b}{\alpha\sigma^2} \tag{7.18}$$

We derive the following interesting insights from these expressions. First, as the risk aversion α or the market price volatility σ^2 increases, the forward hedging quantity h^* increases. Moreover, if the mean market price is equal to the forward price, (i.e., $\mu = b$), the firm will *fully hedge* her production output, (i.e., $x^* = h^*$). If the market price is higher than the forward price, $\mu > b$, then the firm sells some of the output (*partial hedge* $x^* > h^*$). Lastly, if the mean market price is less than the forward price, $\mu < b$, the firm sells more than the production level, hence *speculates* (i.e., $x^* < h^*$).

If the forward price is less than the value obtained in the absence of forward hedging (i.e., $b < \mu - \alpha x^*\sigma^2$), we observe that the forward hedging is done via purchasing from the forward market rather than selling to the market (i.e., $h^* < 0$).

7.2.1.2 Hedging with Other Financial Instruments. Not only forwards/futures trading, but also options are allowed to hedge price risk in this section. Using options creates flexibility in terms of financial instruments and yields interesting managerial insights.

Moschini and Lapan (1992) present a two-stage model of production with risk hedging via futures contracts and options. Main assumptions are as follows:

1. There is no basis risk. Basis risk exists since the location, quality, timing features of the output could differ from the futures contract specifications.
2. Strictly concave utility function is used to capture firm's risk aversion.
3. Convexity of profits in price is captured via quadratic function of price. A symmetric price distribution is assumed to obtain comparative static results.

The model has two stages. In the first stage, the firm commits the quasi-fixed input (capital investment) and determines the open futures and options positions when the output price is uncertain. This is called the *ex ante* problem. In the second stage, the *ex post* problem is choosing the optimal level of output conditional on the futures and options positions taken in the first stage as well as the quasi-fixed input. In the second stage, output price becomes known. The options traded are *straddle options* that are combinations of a call and a put option with equal strike prices. (See, for example, Hull (2003), Chapter 9 for more details.)

Moschini and Lapan (1992) obtain the separation result that the optimal choice of capital investment is not affected by expectations about the stage two price level, but depends on the known futures price. The intuition behind this result is mainly based on the absence of basis risk and production uncertainty as well as the capability to hedge with futures. Thus, the optimal futures hedge is a short position equal to the expected output when the futures prices are unbiased. Moreover, with unbiased futures and options prices, the optimal hedge with options is a *short straddle position*. The key results are as follows:

1. Price volatility raises the lower bound on the optimal straddle hedge.
2. For all risk averse firms, addition of short straddle hedges to the hedging portfolio stochastically dominates portfolios with fewer straddles.
3. *Ex post* elasticity of supply increases the lower bound on the optimal straddle hedge.

In addition, using straddle hedges allows the firm to modify the *ex ante* input level. Addition of short straddles will increase/not affect/decrease the optimal *ex ante* input level, as the shadow price function of the capital investment is convex/linear/concave, assuming quadratic function, symmetrical price distribution, and CARA. Moreover, while short straddles do not change the risk preferences of the firm, they reduce the riskiness of the portfolios, which makes the firms act like risk neutral firms. In summary, the fundamental result of the Moschini and Lapan (1992) model is as follows: If there are fixed input decisions such as capital/capacity investments that are made prior to actual production decisions, then hedging with options in addition to the futures could add value by reducing the riskiness of the portfolios. However, when the production decisions are given *ex ante*, options have no value as the profit is linear in price.

Lapan et al. (1991) study the production and the hedging decisions of a risk averse firm in an expected utility maximization setting with price and basis risks. In their setup, production is also nonstochastic. The firm maximizes the expected utility that has CARA form.

Main results of the paper are similar to the ones noted above. The firm may have unbiased expectations regarding the price distribution and the option premiums. In such a case, options are shown to be redundant hedging instruments. Optimal hedging strategy involves only futures and the amount of futures is determined by the covariance of cash and future prices. When there is no basis risk or CARA utility function is used as well as the cash price is a linear function of the futures price, then separation occurs between hedging and production decisions. As long as the price risk is linear, futures contracts provide a perfect hedge without any options. Nevertheless, straddle options are useful market instruments to hedge the risk due to open speculative position in futures.

EXAMPLE 7.3

Suppose that we have a risk averse firm deciding on the optimal production level, y, given the capital investment made *ex ante*, denoted by z. The price is denoted by p which is resolved *ex post*. Therefore, the *ex post* problem is given by (7.19) where $c(y, z)$ is the variable cost conditional on the capital investment made *ex ante*.

$$g(p, z) \equiv \underset{y}{\text{Max}}\, py - c(y, z) \tag{7.19}$$

In the *ex ante* problem, with random price, the firm has to decide on the optimal decisions for futures and option positions, as well as capital investment, x, s, z respectively. We assume that the firm can trade straddle options on futures that are combination of calls and puts with the same strike price.

The firm's profit is expressed as follows:

$$\tilde{\Pi} = g(\tilde{p}, z) - rz + (f - \tilde{p})x + (t - \tilde{v})s \tag{7.20}$$

In (7.20), r is the price of the capital investment, f is the futures contract price at time 1 for delivery at time 2, t is the price of the straddle and $\tilde{v}$ is the payoff of the straddle which can be expressed as $\tilde{v} = |\tilde{p} - k|$, where k is the strike price of the option. Thus, the *ex ante* problem is to maximize the expected utility of profits by optimally determining the z, x, s. The utility function is strictly concave. Then we have the objective function:

$$\underset{z,x,s}{\text{Max}} E[U(\tilde{\Pi})]$$

The FOCs can be written as follows where the subscripts denote the partial differentiation arguments:

$$\begin{aligned} E[\tilde{U}'[g_z(\tilde{p}, z) - r]] &= 0 \\ E[\tilde{U}'(f - \tilde{p})] &= 0 \\ E[\tilde{U}'(t - \tilde{v})] &= 0 \end{aligned}$$

First, the optimal level of capital investment, z^* can be obtained by using the separation result shown by Holthausen (1979) and Feder, Just and Schmitz (1980). Then, given this value, one can solve for the optimal futures and straddle positions. Assume that futures and straddle prices are unbiased (i.e., $E(\tilde{p}) = f$), and $E(\tilde{v}) = t$. Besides, assume that the profit function $g(p, z)$ is quadratic in price $\tilde{p}$, where $\tilde{p}$ has a symmetric distribution. Given these assumptions, one can show that the optimal futures hedge is a short position equal to the expected production output. That is $x^* = \bar{y}$, the optimal futures hedge is independent of the straddle position. Similarly, one can also show that the optimal straddle hedge is a short position (i.e., $s^* > 0$).

An interesting question is how the straddle options affect the optimal *ex ante* capital investment decision. To be able to show this impact, assume that the utility function has CARA form. Moreover, let us define $\tilde{p} \equiv \bar{p} + \varepsilon$, where $\bar{p} = E(\tilde{p})$ and ε is a zero mean random variable. Suppose the strike price is chosen as $k = \bar{p}$, which means $\tilde{v} = |\varepsilon|$. The FOC for z can be written for the positive realizations of ε as

$$E^{+}\{U'(\Pi(\varepsilon))[g_z(\varepsilon, z) - r] + U'(\Pi(-\varepsilon))[g_z(-\varepsilon, z) - r]\} = 0 \quad (7.21)$$

Using the quadratic feature of $g(.)$ and differentiating with respect to s and incorporating the risk aversion parameter as α we obtain

$$J_s(z, s) = E^{+}\{U'(\Pi(\varepsilon))[g_z(\varepsilon, z) - r + \frac{1}{2}g_{zpp}(z)\varepsilon^2][m - \alpha(t - \varepsilon)]\} \quad (7.22)$$

This holds for any m. Also note that $g_{zpp}(z) = g_{zpp}(p, z)$ due to the quadratic assumption for $g(p, z)$.

Analyzing this expression, it can be shown that as $g_{zpp}(z) \gtreqless 0$, $J_s \gtreqless 0$, thus $\frac{dz^*}{ds} \gtreqless 0$ as $g_{zpp}(z) \gtreqless 0$. This result states that an increase in the short straddle position will increase/not affect/decrease the optimal *ex ante* capital investment as the shadow price of the price function of the capital investment is convex/linear/concave in price.

7.2.2 PRODUCTION PLANNING UNDER MULTIPLE UNCERTAINTIES

7.2.2.1 Price and Basis Risks. To address the joint price and basis risks, Anderson and Danthine (1981) study a hedging problem of a risk averse firm that maximizes mean-variance objective function. Mean-variance objective function is equivalent to expected utility function when the net revenues are normally distributed and the utility function is exponential.

The firm has two decisions to be made at date 0. First one is how much commodity to sell or buy at date 1 at the prevailing cash price. This is the cash position of the firm.[3] Second, the firm has to decide on the optimal futures position at date 0, where offsetting the position is done at date 1. The notion of cross hedging is incorporated into the model. *Cross hedging is defined as taking positions in a related futures markets when the actual commodity of interest does not have its own futures market.* The futures hedge position is shown to be composed of two parts, pure hedge and pure speculation. The covariance between cash and futures prices is crucial in determining the cross hedge possibility. The larger the covariance, the more effective a single cross hedge is in risk reduction. Anderson

[3] Cash position of the firm refers to the real production decision in futures markets terminology. Cash price refers to the spot price.

and Danthine (1981) also examine the impact of futures markets availability on firm's decisions. They show that increasing the list of futures will increase real production given that all of these futures present no speculation incentive.

Example 7.4

Assume that a risk averse firm commits herself in period 0 to sell or buy y amount of commodity in period 1 at price p_1. Production cost is determined by a convex function $c(y)$. In addition, the firm trades futures contracts that may be different in their delivery date, delivery location, type and quality of the commodity. Let f denote the quantity of futures sold in period 0 and $f > 0$ represents sales of futures contracts. This position is closed out in period 1 by an offsetting trade. The futures prices for periods 0 and 1 are given as p_0^f and p_1^f respectively. f and p_t^f are assumed to be column vectors of n futures positions and prices at period t. In period 0, period 1 prices are unknown, thus $\tilde{p_1}$ and $\tilde{p_1}^f$ are random variables with known joint distributions. Period 1 net revenue of the firm can be expressed as

$$\tilde{\Pi} = \tilde{p_1}y - c(y) - (\tilde{p_1}^f - \tilde{p_0}^f)'f \tag{7.23}$$

For the producer of the commodity, $c(y)$ can be considered as the all real production costs including the storage costs.

The firm's mean-variance objective function is shown below in (7.24) where α is the risk aversion coefficient and the firm optimally chooses the cash and the futures contract (y, f) positions.

$$\underset{y,f}{\text{Max}} E\tilde{\Pi} - \frac{1}{2}\alpha \text{Var}\tilde{\Pi} \tag{7.24}$$

The FOCs are given by

$$\begin{aligned} \bar{p_1} - c'(y) - \alpha(y\sigma_{00} - \Sigma_{01}f) = 0 \\ (p_0^f - \bar{p_1}^f) - \alpha(\Sigma_{11}f - y\Sigma_{10}) = 0 \end{aligned} \tag{7.25}$$

The covariance matrix for the joint distribution of the cash and the futures prices is given by

$$\begin{bmatrix} \sigma_{00} & \Sigma_{01} \\ \Sigma_{10} & \Sigma_{11} \end{bmatrix}$$

Note that $\sigma_{00} = \text{Var}(\tilde{p_1})$ and Σ_{11} is $n \times n$ matrix representing the covariance matrix for the prices of the different futures contracts. Also note that if $c''(y) \geq 0$ and Σ is non-negative definite, the mean-variance objective function is concave in (y, f), hence the FOCs are necessary and sufficient for a maximum.

One can then compute the optimal cash and futures positions. Given the optimal cash position y, the optimal futures position is computed as

$$f = \frac{1}{\alpha}\Sigma_{11}^{-1}(p_0^f - \bar{p_1}^f) + y\Sigma_{11}^{-1}\Sigma_{10} \quad (7.26)$$

Note that there is a unique solution for the futures if and only if Σ_{11} is nonsingular.

There are two cases to be analyzed here. First one is when $y = 0$ (i.e., *pure speculation*). In this case, the optimal futures position will be expressed as $f = \frac{1}{\alpha}\Sigma_{11}^{-1}(p_0^f - \bar{p_1}^f)$. Therefore, the optimal position depends on the risk aversion coefficient, the covariance matrix, and the vector of expected returns. Second case is where there is hedging (i.e., $y \neq 0$). In this case, one can divide the optimal futures position into two parts, one being a *pure hedge*, which is $y\Sigma_{11}^{-1}\Sigma_{10}$, and the other a pure speculation, the remaining term.

Considering a single cross hedge, $n = 1$, the pure hedge term becomes $y\sigma_{01}/\sigma_{11}$. Note that the sign of the pure hedge depends on the sign of y and σ_{01}. Hedging with futures is usually conducted for the same commodity that is being produced (i.e., $\sigma_{01} > 0$). However, there may be hedging for a commodity where the same type of futures may not exist in the market, which necessitates the cross hedging for a related commodity. For example, a firm may hedge the stainless steel price risk by trading nickel futures. In those cases, $\sigma_{01} < 0$ holds. Further, one can express the pure hedge as shown below where ρ is the correlation coefficient between futures and cash prices.

$$|\sigma_{01}/\sigma_{11}| = |\rho|\sqrt{\sigma_{00}/\sigma_{11}} \quad (7.27)$$

Thus, we conclude that as $|\sigma_{01}/\sigma_{11}|$ increases, a single cross hedge becomes more effective in risk reduction.

On the other hand, one can compute the optimal cash position as follows: Using the optimal futures position given in the equation (7.27) above and incorporating it to the FOC for the optimal cash position, we obtain the necessary condition for the optimal y given that the f is dynamically adjusted. We obtain

$$\bar{p_1} + \Sigma_{01}\Sigma_{11}^{-1}(p_0^f - \bar{p_1}^f) - y\alpha(\sigma_{00} - \Sigma_{01}\Sigma_{11}^{-1}\Sigma_{10}) - c'(y) = 0 \quad (7.28)$$

Instead of multiple hedges, one can also write this condition for a single cross hedge as

$$(1-\beta)\bar{p_1} + \beta(p_0^f - \bar{B}) - y\alpha\sigma_{00}(1 - R^2) - c'(y) = 0, \quad (7.29)$$

Note that R^2 is the theoretical multiple correlation coefficient squared, $\beta = \sigma_{01}/\sigma_{11}$ and $\bar{B} = \bar{p_1}^f - \bar{p_1}$ being the expected value of the basis, the difference between the cash and the futures price. This expression can be interpreted as follows: The first two terms can be considered as the *planning price*, which is a weighted average of the expected cash price and all current and next period futures prices. Third term is the *risk premium* that depends on the size of the cash position, the risk aversion coefficient, the cash price variance, and the proportion of variability that cannot be eliminated via optimal hedge in the futures. The fourth term is the marginal cost of production.

7.2.2.2 Price and Basis Risks with Costly Hedging. Frechette (2001) studies the optimal hedging portfolio with options and futures in an expected utility maximization setting where hedging is costly. Besides, price and basis risks exist. In the previous work by Lapan, Moschini and Hanson (1991), hedging was assumed to be costless, which led to optimal portfolios where hedging with options are not used. In contrast, Frechette (2001) shows that costs of hedging makes the optimal hedge ratios with options nonzero.

When hedging costs are considered, optimal hedging will consist of a mix of options and futures. Hedgers should carefully consider the costs and benefits of hedging. Overhedging need not be speculative. It may occur in cases where marginal cost of futures trading increases the hedge ratio. As the risk aversion rises, overhedged position declines. Options hedging is shown to be more sensitive to costs than futures hedging. If transaction, learning, and management costs related to options hedging can be reduced, overall hedge ratios can be changed. Thus, options hedging could be thought as a luxury good whereas futures hedging is a necessary good. Futures and options could be considered as imperfect substitutes. Risk aversion plays a critical role in hedging behavior. Low risk aversion makes the hedging insignificant. However, as risk aversion increases, hedgers are almost indifferent between optimal futures-only strategy and optimal options-only strategy. In this case, optimal futures-only hedging strategy involves a lower hedge ratio than the optimal options-only hedging strategy.

EXAMPLE 7.5

Assume that the risk averse firm produces and hedges the price risk using futures contracts and put options in order to maximize the expected utility of profit. The production cost is strictly convex and given by $c(y)$ where y is the production quantity. Let x be the quantity of futures contracts sold and z the amount of put options sold. Let f and p be the futures prices at the start and end of the period, respectively. Moreover, we assume that b is the cash/spot price, including the basis risk, at the end of the period, r is the put option price and the v is the terminal value of the put option with k being the strike price. The end of period profit can be expressed as follows:

$$\tilde{\Pi} = \tilde{b}y - c(y) + (f - \tilde{p})x + (r - \tilde{v})z \tag{7.30}$$

Note that $\tilde{v}$ is the terminal value of a put option, computed as

$$\begin{aligned} \tilde{v} &= 0, && \text{if } p \geq k \\ \tilde{v} &= k - \tilde{p}, && \text{if } p < k \end{aligned} \tag{7.31}$$

The firm wants to maximize the expected utility by choosing y, x, z optimally. Thus the objective function can be written as

$$\underset{y,x,z}{\text{Max}} E[U(\tilde{\Pi})] \tag{7.32}$$

The relationship between the cash/spot price and the futures prices needs to be specified. Following the analysis of Lapan et al. (1991), let us assume that the cash/spot price is a linear function of the futures price, thus

$$\tilde{b} = A + \beta\tilde{p} + \tilde{\theta}, \tag{7.33}$$

Both $\tilde{p}$ and $\tilde{\theta}$ are assumed to be independently distributed. Note that $\tilde{\theta}$ denotes the orthogonal basis risk. Also note that $E(\tilde{\theta}) = 0$. Using this relationship, computing the FOCs for y, x, z respectively and equating to zero, we obtain

$$\begin{aligned} E[\tilde{U}'(A + \beta\tilde{p} + \tilde{\theta} - c')] &= 0 \\ E[\tilde{U}'(f - \tilde{p})] &= 0 \\ E[\tilde{U}'(r - \tilde{v})] &= 0 \end{aligned} \tag{7.34}$$

Assume that the utility function has CARA representation with risk aversion parameter α. Then, one can write the FOC for y using CARA, where $h(\theta)$ is the probability density function of $\tilde{\theta}$, as follows:

$$\frac{\int \theta e^{-\alpha\theta y} h(\theta) d\theta}{\int e^{-\alpha\theta y} h(\theta) d\theta} = c'(y) - A - \beta f \tag{7.35}$$

It is clear to observe that the distribution of $\tilde{p}$ does not affect the optimal production output y^*. Thus, the separation result aforementioned is shown to hold under the assumptions of CARA and the linear relationship between the futures and the cash/spot prices.

Moreover, in the absence of CARA, yet assuming unbiased futures prices (i.e., $E(\tilde{p}) = f$) as well as option prices (i.e., $E(\tilde{v}) = r$), the FOCs can be simplified by writing as the covariances

$$\begin{aligned} \mathrm{Cov}(\tilde{U}', \tilde{\theta})/E\tilde{U}' &= c'(y^*) - A - \beta f \\ \mathrm{Cov}(\tilde{U}', \tilde{p}) &= 0 \\ \mathrm{Cov}(\tilde{U}', \tilde{v}) &= 0 \end{aligned} \tag{7.36}$$

If there is no basis risk (i.e., $\tilde{\theta} = 0$), then the first FOC given above becomes $c'(y^*) = A + \beta f$, which means that separation of production and hedging occurs as optimal production output y^* is independent of the distribution of $\tilde{p}$. Second, when $\beta = 1$, meaning the basis $(\tilde{b} - \tilde{p})$ and the futures price $\tilde{p}$ are independent, the optimal hedge is the full hedge in the futures market, yet separation is not necessarily obtained. Third, using these FOCs, one can verify that the optimal hedging with futures contracts and the options are given as $x^* = \beta y$ and $z^* = 0$. Then, the optimal production output has to satisfy $c'(y^*) < A + \beta f$ by previously mentioned results of Sandmo (1971), where the random price is $A + \beta f + \tilde{\theta}$. This result demonstrates that in the presence of futures contracts, options provide no value in the optimal hedging portfolio as the hedgeable risk is linear in $\tilde{p}$ and the futures yield a linear payoff in $\tilde{p}$.

7.2.2.3 Price, Basis, and Yield Risks. Moschini and Lapan (1995) address the optimal hedging strategy of a risk averse firm in the presence of futures price, basis and production risks. In the earlier work mentioned above by Lapan et al. (1991), production is assumed to be nonstochastic. In this paper, however, they assume the existence of production uncertainty, given as *yield risk*, in addition to price and basis risks. These risks are modeled as multivariate normal distributions. They mainly show that production risk introduces the use of options on futures in the optimal hedging portfolio.

The model operates in a single period where all input and the hedging decisions are taken at once by the producer. Straddle options are used, each composed of a call and put option with the same strike price. The firm has CARA utility function. Under the additional risks:

- If the futures price is perceived unbiased, then the optimal futures position is computed regardless of the number of the straddles. However, price bias necessitates that the optimal hedging should be determined simultaneously with the straddle position.
- If futures and option prices are perceived unbiased and if the yield random variable is not positively correlated with the futures price, then the optimal hedging strategy consists of short future contracts and long straddles. The interaction between yield and future price uncertainties creates the nonlinearity of the profit function, which leads to net straddle position. In essence, straddles are used to hedge the speculative future positions.

- Futures and options hedging availability induces the firm to expand production output given that the risk aversion level is not too high.

Moschini and Lapan (1995) apply their methods for soybean production and hedging on Iowa cross-sectional time series data set. Interesting managerial insights are derived. First, sensitivity analysis on the basis and yield risks is conducted to observe how the optimal hedge ratios change. Basis risk increases the futures and absolute straddle hedge ratios. However, yield risk reduces the futures hedge ratio, yet increases the straddle hedge ratio. Thus, both risks tend to reinforce each other for the straddle hedge, but weaken each other's impact for the futures hedge. Impact of the perceived bias for the futures and the straddle price is also analyzed. Perceived bias for the futures price increases the futures hedge ratios in a nonlinear fashion. In addition, as the futures price bias changes in either direction, producers increase the long straddle positions. On the other hand, straddle price bias does not affect the futures hedge ratios, but it reduces the long straddle positions and leads towards short straddles. Furthermore, all of these results are shown to be sensitive to the risk aversion level.

EXAMPLE 7.6

Assume that the risk averse producer decides on the production input that yields a random output in the end, which is unknown at the time of the decision. Let $q = f(X)$ denote the scale of production, where X is the vector of inputs. The total output can be written as $\tilde{Y} = \tilde{y}q$, and the cost function is denoted as $c(q)$ where $\tilde{y}$ is a random variable with support $[0, 1]$. This uncertainty for the production is named as the yield risk.

Let the producer earns p_2 per unit of output sold in the market. Let p_1 denote the price at which the hedging is settled. Difference between these two prices creates the basis risk. The firm can hedge the price, basis, and yield risks by trading futures contracts at price p_f and straddle options with given strike price k and option price r. Thus, the payoff for a straddle option can be computed as $r - |p_1 - k|$. If the producer sells futures contracts and sells straddle options, the profit function can be written as

$$\tilde{\Pi} = \tilde{p_2}\tilde{y}q - c(q) + (p_f - \tilde{p_1})f + (r - |\tilde{p_1} - k|)z \tag{7.37}$$

The Variables q, f, z are the amounts of the quantity produced, futures contracts, and straddle options sold respectively.

Assuming expected utility maximization, the FOCs for the (q, f, z) can be expressed as follows:

$$\begin{aligned} E[\tilde{U}'(\tilde{p_2}\tilde{y} - c'(q))] &= 0 \\ E[\tilde{U}'(p_f - \tilde{p_1})] &= 0 \\ E[\tilde{U}'(r - |\tilde{p_1} - k|)] &= 0 \end{aligned} \tag{7.38}$$

In order to model the correlation between the cash price, yield, and futures prices, let us define the random variables as follows:

$$\tilde{p_1} \equiv \bar{p_1} + \tilde{\varepsilon_1}$$
$$\tilde{p_2} \equiv \bar{p_2} + \tilde{\varepsilon_2} \qquad (7.39)$$
$$\tilde{y} \equiv \bar{y} + \tilde{\varepsilon_3}$$

In this definition, ε_i's denotes zero mean random variables and the $\bar{p_1}, \bar{p_2}, \bar{y}$ are conditional means of the random variables. Moreover, let $\tilde{\eta_i} \equiv \tilde{\varepsilon_i} - \beta_i \tilde{\varepsilon_1}$ for $i = 2, 3$, where β_i is a scalar such that $E[\tilde{\eta_i}\tilde{\varepsilon_1}] = 0$. Let the strike price be selected such that $k = \bar{p_1}$. Under these assumptions, the FOCs can be reexpressed as follows:

$$E[\tilde{U}'(\bar{p_2}\bar{y} - c'(q) + \gamma_1\tilde{\varepsilon_1} + \frac{\gamma_2\tilde{\varepsilon_1}^2}{2} + \delta'\tilde{\eta} + \frac{\tilde{\eta}'S\tilde{\eta}}{2})] = 0$$
$$E[\tilde{U}'(\theta - \tilde{\varepsilon_1})] = 0 \qquad (7.40)$$
$$E[\tilde{U}'(r - |\tilde{\varepsilon_1}|)] = 0$$

where $\theta = p_f - \bar{p_1}$ and
$$M \equiv \begin{bmatrix} \bar{y} \\ \bar{p_2} \end{bmatrix}, S \equiv \begin{bmatrix} 0 & 1 \\ 1 & 0 \end{bmatrix}, \eta \equiv \begin{bmatrix} \eta_2 \\ \eta_3 \end{bmatrix}, \beta \equiv \begin{bmatrix} \beta_2 \\ \beta_3 \end{bmatrix}, \gamma_1 \equiv M'\beta, \gamma_2 \equiv \beta' S\beta,$$
and $\delta' \equiv (M' + \tilde{\varepsilon_1}\beta' S)$

Assuming that the risk averse producer has CARA and the random variables ε_i are jointly normally distributed, then using the FOCs described above, one can obtain the optimal futures and options hedging values. After lengthy computations, it is shown by Moschini and Lapan (1995) that for biased futures prices, where $\theta \neq 0$, the optimal futures position cannot be decided independently of the straddle position. In the presence of unbiased futures prices only, ($\theta = 0$), futures contracts position can be decided independently of the straddle position. On the other hand, it is also shown that in the case of unbiased futures and option prices, if the yield has nonpositive correlation with the futures price, the optimal portfolio contains buying straddles in addition to selling futures.

7.3 Multiperiod Models

In this section, we review models that address the integrated production and hedging decisions of a risk averse firm in multiperiod settings. The firm maximizes the expected utility or mean-variance objective function. Later, we present our work for a risk neutral producer that is exposed to demand and price risks who optimally decides on the production and spot market trading.

7.3.1 PRODUCTION PLANNING UNDER PRICE UNCERTAINTY

We review two models in the general utility maximization setting, which helps in deciding on the optimal production level in the presence of forwards/futures. In Antonovitz and Nelson (1988), a risk averse firm maximizes the expected utility by trading the forward/futures contracts and spot sales in a two period setting. Main results can be listed as follows:

1. Production and trading in the market decisions can be decoupled in the unrestricted trading case (separation result).
2. The firm produces the optimal production level where marginal cost of production is equal to the forward price. The optimal production level is independent of the risk aversion and the probability distribution of the market price.
3. **(a)** If the expected spot price is less than the forward price, selling forwards is optimal, yet this case is not sustainable due to arbitrage opportunities as the forward price converges to the expected spot price. The firm also purchases futures contracts.
   (b) When the expected spot price is equal to the forward price, selling forwards is the riskless alternative.
   (c) If the expected spot price is higher than the forward price, firm shorts futures position and also decides on the hedge ratio by examining the covariance between the futures and spot prices and the variance of the futures price. If the variance of the futures price exceeds the covariance (futures, spot prices), the firm maintains an open spot position.
4. When the firm is restricted to only hedging and not speculation (i.e., no short selling is allowed for forwards/futures and spot market), then production and trading activities can be decoupled only under certain conditions. First, the firm compares the forward price with the spot price to determine the futures contracts trading option. Forward contract price being higher than the spot price leads the firm to fully sell via forward contracts, the riskless option. On the other hand, higher spot prices direct the firm to compare the covariance (futures, spot prices) with the variance of futures prices. Again if and only if the variance of futures prices exceed the covariance (futures, spot prices), a net short spot position is taken.

   It can be shown that the forward contracts can be excluded from the feasible decision set when:

   - The marginal cost of producing the futures position is higher than the forward price, future commitment is positive and open spot market position is zero.
   - The marginal cost of producing the futures and the open spot market positions are higher than the forward price, future commitment is positive and open spot market position is positive.

In sum, Antonovitz and Nelson (1988) show that the forward contract price is used as a riskless benchmark when determining the optimal futures and spot market positions. It also allows the decoupling of production and trading decisions. Futures contracts on the other hand are used mainly to hedge risk without changing the profitability of the firm.

7.3.2 PRODUCTION PLANNING UNDER MULTIPLE UNCERTAINTIES

7.3.2.1 Price and Yield Risks. Anderson and Danthine (1983) and Bowden (1995) study a multiperiod model of hedging with futures for risk averse hedgers. Anderson and Danthine (1983) consider the expected concave utility maximization model where there are price and yield risks, but no basis risk. In their model, there are multiple trading dates where at the first date, inputs are procured at exogenous input price and the output is sold at prevailing cash price at the last trading date. On the other hand, futures contracts can be traded at each date during the horizon. They use backward recursion to obtain the optimal positions for cash and the futures contracts. Without providing the mathematical details, we list the main results of the paper below:

- Cash and futures positions can be separately decided given that there is neither yield nor basis risk. This is the aforementioned separation result.
- The hedger's optimal futures position can be decomposed into a pure hedge and a pure speculation. The pure hedge increases in absolute value as the delivery date approaches when the interest rate is nonstochastic.
- If and only if the futures price is expected to increase (decrease), the pure speculation part of the futures is long (short).
- In the presence of production uncertainty, separation result holds under certain conditions. For example, the pure hedge of the futures at time 2 is equal to the time 2 expectation of time 3 output. Similarly, the pure hedge at time 1 is equal to the time 1 expectation of time 3 output discounted at time 2 interest rate. Thus, the hedge evolves dynamically over time but in a fairly predictable manner.

7.3.3 PRODUCTION PLANNING IN THE PRESENCE OF PRICE AND DEMAND RISK IN A CONTINUOUS TIME SETTING

In this section, we present a model for a risk neutral commodity producer who has to satisfy stochastic demand of a buyer for the duration of the longterm contract at fixed price using her own production as well as spot market. Therefore, the firm is exposed to both demand and spot market price risks. The key differences with previous work is the treatment of time and risk neutrality of the firm. We shall see that the continuous time formulation makes the analysis somewhat simpler.

Our model minimizes the expected costs associated with this operation. We allow the producer to purchase from the spot market even when she can use her own capacity to satisfy the demand. In that sense, spot market trading is both for dumping and procuring commodities regardless of capacity usage. It may be profitable to operate this way as will be seen in our results. The capacity of the producer is allocated exclusively for the buyer and is fixed during the horizon. Uncertainties in demand and spot price, together with other factors, affect the value of spot market for the producer.

The spot market price is modelled as a Geometric Brownian motion (GBM). (Li and Kouvelis 1999, Brennan and Schwartz 1985). Our model can also incorporate the mean-reverting characteristic for the spot price. It is important to note that mean-reversion behavior in spot prices is empirically observed for some commodities such as copper, oil, and gold. This phenomenon is well documented in finance literature. See, for example, Schwartz (1997) for a detailed discussion on behavior of commodity spot prices under one and multifactor models that consider the mean reversion in spot price, convenience yield, and interest rates. We think, using GBM to model the spot market prices is appropriate for our purposes to simplify the exposition. On the other hand, we model the cumulative demand observed by the producer as an Arithmetic Brownian motion (ABM). We use ABM for modeling demand since the variable with ABM has Normal distribution whereas GBM implies a Lognormal distribution for the random variable, which is more appropriate for modeling the spot market price.

The two stochastic processes that drive the model are given below:

1. *Cumulative demand process:*

$$dD_t = a_d dt + \rho \sigma_d dW_1 + \sqrt{1-\rho^2}\sigma_d dW_2 \tag{7.41}$$

2. *Spot market price process:*

$$dS_t = b_s S_t dt + \sigma_s S_t dW_1 \tag{7.42}$$

In this formulation, a_d denotes the constant drift of the cumulative demand process, whereas b_s is the constant drift for the spot market price, ρ denotes the correlation coefficient between the cumulative demand and the spot market price processes, σ_d and σ_s denote the volatilities of the cumulative demand and the spot market price respectively, and W_1 and W_2 are independent standard Brownian motions (Wiener processes).

We denote the spot market price as S_t. We model the spot market price and the cumulative demand processes as 2-dimensional Itô process.[4] With this 2-dimensional Itô process, we can model the correlation between the spot market price and the demand. Both positive and negative correlation structures can be modelled. For instance, when demand and spot price are independent, i.e., $\rho = 0$,

[4] See for example Oksendal (2005, p. 48) for a discussion on multidimensional Itô processes. This approach is used frequently to model multiple uncertainties in a compact and efficient way.

then W_1 disappears, thus the two stochastic processes become independent of each other. If the spot price and demand are perfectly positively correlated, i.e., $\rho = 1$, then the diffusion of the inventory level is governed by the Wiener process of the spot market price, W_1, as the second Wiener process, W_2, disappears.

By using the 2-dimensional Itô process, we can model the inventory level at the warehouse of the producer as a stochastic process as follows. Every period, the producer produces a certain amount of commodity, denoted by u_t, and trades in the spot market (both buying and selling are allowed) the amount v_t. Thus, the instantaneous change in the inventory level is given by the following differential equation:

$$\begin{aligned} dI_t &= (u_t + v_t)dt - dD_t. \\ &= (-a_d + u_t + v_t)dt - \rho\sigma_d dW_1 - \sqrt{1-\rho^2}\sigma_d dW_2 \end{aligned} \tag{7.43}$$

Note that the drift of the inventory process is affected by three variables, the negative drift of the cumulative demand, $-a_d$, the quantity produced, u_t, and the quantity traded in the spot market, v_t. The diffusion is due to both the spot price and cumulative demand processes due to 2-dimensional Itô process as described above.

Now, we state our stochastic control problem as follows:

$$\begin{aligned} V(S) &= \min_{u_t, v_t \in U} E\left(\int_0^\infty e^{-rt}(h_t I_t^+ + b_t I_t^- + c_t u_t + S_t v_t)dt\right) \\ s.t.& \\ 0 &\le u_t \le K_t, \quad \forall t \\ dD_t &= a_d dt + \rho\sigma_d dW_1 + \sqrt{1-\rho^2}\sigma_d dW_2 \\ dS_t &= b_s S_t dt + \sigma_s S_t dW_1 \\ dI_t &= (-a_d + u_t + v_t)dt - \rho\sigma_d dW_1 - \sqrt{1-\rho^2}\sigma_d dW_2 \end{aligned}$$

In this formulation, the producer minimizes the expected total costs, that are composed of the production, inventory holding, backordering and spot trading costs in the infinite horizon. Note that X^- denotes the negative part of $X, \forall X \in \mathbb{R}$. h_t denotes per unit inventory holding cost, b_t per unit backorder cost, and c_t denotes per unit production cost in period t, $\forall t$. Since the producer first satisfies the random demand of the buyer, irrelevant of what happens in the spot market price, maximizing the expected profit is equivalent to minimizing total expected costs given that the per unit sales revenue is constant. Also note that the producer is allowed to sell commodity to the spot market, in which case the total costs are reduced. There is a finite production capacity, denoted by K_t and r denotes the risk-free interest rate. Both control variables (u_t, v_t) are bounded. We assume that $-\infty < v_t < \infty$ should hold. Moreover, if $v_t < 0$, then $-v_t \le I_t + K_t$ also needs to hold. In this model, we assume that the production level can be changed without incurring extra costs (i.e., no production friction exists).

We can write out the Hamilton-Jacobi-Bellman (HJB) equation for this model as follows:

$$rV = \min_{u_t, v_t \in U} [(h_t I_t^+ + b_t I_t^- + c_t u_t + S_t v_t) + V_I(-a_d + u_t + v_t) + V_S(b_{s)} S_t) + \frac{1}{2}\sigma_{d)}^2 V_{II} + \frac{1}{2}\sigma_s^2 S_t^2 V_{SS}] \quad (7.44)$$

Proposition 7.1 states the optimal control strategy for the risk neutral commodity producer.

Proposition 7.1 *When the producer is able to produce the amount u_t, $0 \leq u_t \leq K_t$ and trade (buy/sell) in the spot market the amount v_t, in the absence of production friction, the optimal values for the controls are given as follows:*

$V_I + S_t < 0$ can never happen, if it were, $v_t^* = \infty$, which is physically impossible. Thus, $V_I + S_t > 0$ holds. Then, we have the following cases:

Case 1: $V_I + c_t < 0$, then $u_t^* = K_t$, $v_t^* = -(I_t + K_t)$.

Case 2.1: $V_I + c_t > 0$ and $c_t > S_t$, then $u_t^* = 0$, $v_t^* = -I_t$.

Case 2.2: $V_I + c_t > 0$ and $c_t < S_t$, then $u_t^* = K_t$, $v_t^* = -(I_t + K_t)$.

Proof: We observe that the state transition dynamics and the reward function are linear in the control variables. In addition, the control variables are bounded. Therefore, the optimal control policy has a bang-bang type structure (Sethi and Thompson 2000, Chapter 13). In other words, it is optimal to set the control variables either at the lower or the upper boundary values. We can explicitly determine the optimal control policy by using Karush-Kuhn-Tucker conditions. We have two cases depending on the sign of I_t.

Note that $V_I + S_t < 0$ can never happen. If it were to happen, we could buy an infinite amount of commodity (i.e., $v_t^* = \infty$) from the spot market, which is physically impossible. Therefore, $V_I + S_t > 0$ has to hold for all cases. Then, we have two cases to consider depending on the comparative values of the production cost c_t and the prevailing spot market price S_t. We obtain the optimal control values as given in the theorem QED.

This optimal control strategy implies the following: We produce up to the capacity or cease production and buy or sell in the spot market depending on the relative magnitudes of inventory level and the production capacity. Intuitively, Case 1 implies the following. The value function (total costs) has an inclination to decrease with respect to the production cost. Therefore, the firm is better off producing up to the capacity and decide on selling/buying in the spot market depending on the level of inventory using whatever is available. In other words, if the firm has backlogs, then she may need to buy from the spot market in case the production capacity is not sufficient. On the other hand, if the firm has positive inventory-on-hand it dumps to the spot market this inventory-on-hand as well as the recently produced. Case 2.2 has a similar intuition. In that case, even though the value function has an inclination to increase with respect to the production cost, the spot market price is higher than the unit production cost.

Hence, the firm also produces up to the capacity level and then sells/buys based on the sum of the capacity and the inventory level. Lastly, Case 2.1 depicts the situation where value function tends to increase with respect to the production cost and the prevailing spot market price is favorable than the unit production cost. The firm then produces nothing and goes to the spot market to *buy (sell)* in case there is *inventory backlogs (positive inventory-on-hand).*

7.3.3.1 Numerical Study. Since the HJB equation (7.44) is not solvable analytically due to the partial differentiation of the value function involved, we resort to numerical computation. In this area of research, abundant amount of work has been done to solve ODEs and PDEs for the stochastic control problems that arise in finance, economics, and supply chain management literature. Kushner (1977) and Kushner and Dupuis (2001) extensively treat the subject of solving elliptic PDEs under Markov discretization process of the continuous problems under variety of conditions, finite, infinite horizons, discounted, undiscounted, and soon. One significant application arena of solving the PDEs relates to pricing derivatives, hence researchers in finance proposed methods including the widely celebrated Binomial, Trinomial, and the Multinomial approximation methods. For an overview of different methodologies, please refer to a comprehensive treatment by Hull (2003). Kamrad and Ritchken (1991) propose a Multinomial approximation that is shown to be computationally efficient for valuing contingent claims that have multiple sources of uncertainty. Kamrad and Ritchken (1991) use a derivative-free method based on grid search, which is somewhat different than our approach.

To solve our stochastic control problem, we use the cubic spline collocation method, which can resourcefully solve elliptic partial differential equations. It is shown by Hadjidimos et al. (1999) that cubic spline collocation method is both efficient and accurate with an easy implementation process. According to Hadjidimos, Houstis, Rice and Vavalis (1999), this method combines the features of finite element and finite difference discretization schemes.[5]

On the other hand, our model has a bang-bang control structure, therefore it exhibits kinks in the value function, which destroys the smoothness of the function and creates discontinuities. This feature further strengthens the use of cubic spline collocation method instead of using other functional approximations, such as Chebychev polynomial or linear spline interpolations. Furthermore, approximation errors for various functions in Miranda and Fackler (2002) clearly suggest that the cubic spline collocation method should be preferred for functions that show high degree of curvature or discontinuous derivatives.

Alternatively, one can also use the finite difference approximation, which transforms the continuous time model into a Markov chain on a finite state space.

[5] It is also applicable to any type of PDE that can be written in a tensor product form. Tensor algebra is necessary when multivariate functions are approximated as we require in this process. See Hadjidimos et al. (1999) and references therein on the cubic spline collocation method. In another related work, Tsompanopoulou and Vavalis (1998) formulate and analyze the alternating direction implicit schemes for the cubic spline collocation method, which is also found to perform quite well.

TABLE 7.1 Parameter Values for the Computational Study

Parameter	Base Case	Experiments
Drift of the cumulative demand ($a_{d(t)}$)	10 ton	(20, 30, 50)
Drift of spot market price ($b_{s(t)}$)	10 ton	(20, 30, 50)
Unit inventory holding cost(h)	\$1/ton	(2, 3, 5, 10)
Unit production cost (c)	\$1/ton	(2, 5, 10, 25)
Unit backordering cost (b)	\$1/ton	(0.5, 2, 5, 10)
Correlation coefficient (ρ)	0	(−1, 0, 1)
Volatility of the spot market price ($\sigma_{s(t)}$)	0.20	(0.01, 0.40, 0.50)
Volatility of the cumulative demand ($\sigma_{d(t)}$)	0.10	(0.20, 0.30, 0.50)
Production capacity level (K)	50 ton	(10, 100, 300)
Risk-free interest rate (r)	0.01	(0.10, 0.20, 0.30)

This finite state space is a *disretization* of the original problem's state space.[6] Since our model has two control and two state variables, we think that the cubic spline collocation method is much more efficient and easy to work with.[7]

While computing the PDE numerically, we aim to answer the following questions. What is the value of additional production capacity when the spot market price changes? Under which spot market conditions is the additional capacity most valuable? How do the cost parameters (i.e., inventory holding, production, back-ordering) impact the total production costs and the value of spot market trading? What is the impact of spot price volatility on the optimal control policy?

All cases in the numerical study are solved numerically using MATLAB. For the functional approximation, the grid size has been chosen under the best available computer memory. The approximation errors show that the cubic spline collocation method yields reasonably good solutions. The stochastic control solver code was developed based on the work by Miranda and Fackler (2002).

We normalize a few cost parameters to value 1, such as unit inventory holding cost, h, unit production cost, c, and unit backordering cost, b. We also assume that these parameters are time-invariant. Parameter values for which the numerical computations are performed are given in Table 7.1.

We assume that the cumulative demand is less variable than the spot market price. The volatility of the spot market price is the most critical uncertainty for the producer in order to value the long-term contract. We approximately solve the HJB equation by the cubic spline collocation method via discretization of the

[6] See Kushner and Dupuis (2001) for a complete treatment on this topic.

[7] Besides, there are some advantages while coding it with MATLAB since functional approximations with matrix operations can easily be managed in this environment. Thus, our numerical experiments utilize the cubic spline collocation method to approximate the elliptic PDE, which eventually solves the infinite horizon stochastic control problem.

state space. The value function is approximated at every node of the constructed 2-dimensional grid.[8]

7.3.3.2 Impact of Production Capacity. In this experiment, we examine the impact of production capacity on the value function and the optimal control policy. We assume that increasing this capacity is costless as any upward/downward adjustment is freely possible. A more elaborate model would entail the capacity-related costs, which is overlooked in this model. However, capacity adjustment costs, whenever they are irreversible, might cause a hysteresis zone as documented in previous capacity management literature.[9]

We obtain insights on the value of additional capacity under different spot market conditions. We perform our tests with capacity levels $K = (10, 50, 100, 300)$. Results obtained are as follows. The value of additional capacity diminishes as the spot market prices attain lower values. This result is intuitive because as the spot price drops, the production cost becomes relatively more expensive and thus full capacity is less likely to be used. Thus the value of additional capacity diminishes. We observe that the value function obtains the same values for the following pairs of the state variables:

- Low spot price–high inventory level (positive stock-on-hand)
- High spot price–low inventory level (backlogs)

and

- Low spot price–low inventory level (backlogs)
- High spot price–high inventory level (positive stock-on-hand).

When the value function obtains higher costs, the latter two pairs in the list above are realized. Thus, it is advisable that the producer should prefer having backlogs rather than positive stock-on-hand when the spot price is high, as in the second case listed above, and act vice versa. Simply, when the spot prices are high (low), it is more profitable to sell (buy) in the spot market and incur backordering (holding) cost than holding (backordering) the inventory to fulfill the long-term contract demand. Note that the above result is viable when the production capacity is low. A similar situation is observed in the next experiment when the production cost varies.

As the capacity level increases, the preference should go to purchasing from the spot market when the spot price is low than selling in the spot market when the spot price is high. This is due to the saddle shape of the value function, seen in Figure 7.1, that becomes asymmetric as the capacity level increases. In essence, the pair, high spot price–low inventory level, becomes more costly than

[8] Details of the computation are available from the authors upon request.

[9] See Van-Mieghem (2003) for an excellent survey of the field and Eberly and Van-Mieghem (1997) for a thorough analysis of the hysteresis zone and its implications.

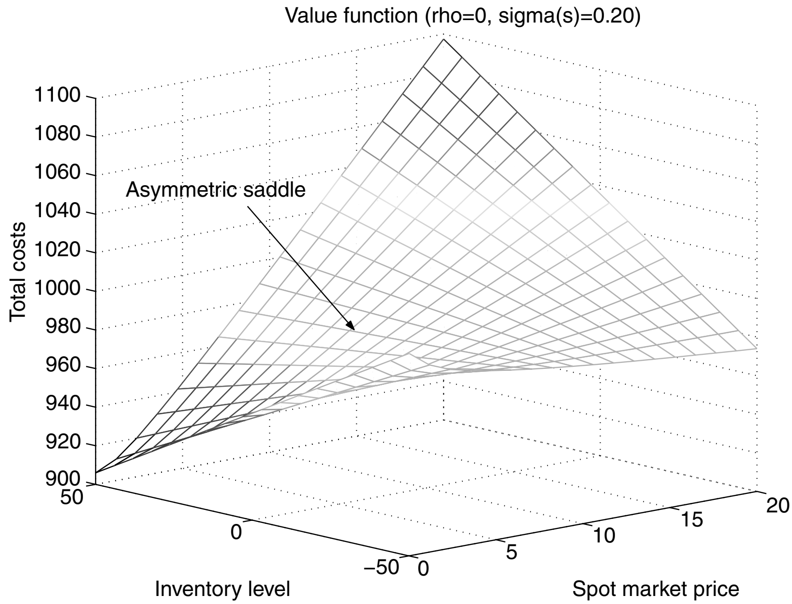


FIGURE 7.1 Asymmetric saddle shape of the value function.

the pair, low spot price–high inventory level, which creates the asymmetry in the structure. Hence, our managerial insight suggests that: Additional capacity favors purchasing more from the spot market when the spot price is low than selling to the spot market when the spot price is high.

This may sound counter-intuitive, yet is a result of the nonlinearity of the value function. We will show that this rule holds under variety of conditions, thus is more general. When the capacity is high we observe that the small changes in the spot price cause dramatic changes in the inventory level. Therefore, the inventory level needs to be adjusted much more sensitively based on the changes in the spot market price. Thus, as another managerial insight, we conclude that: As the production capacity increases, the changes in the spot market price needs to be carefully monitored as it has a stronger effect on the inventory of the producer.

This useful insight holds for the whole range of spot prices except when the prices are very low. As to the changes in the control policy, we can conclude that the control policy preserves the same structure, yet the values traded in the spot market and the commodity produced is scaled according to the changes in the capacity. In the low capacity case, the producer is more willing to purchase from the spot market, whereas in the high capacity case she is more likely to sell to the spot market under the same set of spot price conditions. Besides, the producer with high capacity tends to carry more inventory than the producer with low capacity and thus has more incentive to sell to the spot market to lower the inventory. Besides, as production capacity increases, so do the production quantities.

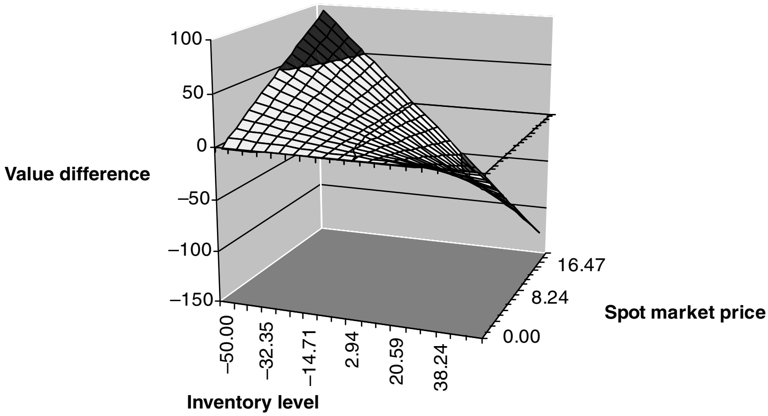


FIGURE 7.2 Value of spot market trading: Low (tight) capacity, ($K = 10$).

Next, we examine the value of spot market trading. One of the important levers for valuing the spot market trading is the capacity available at the commodity producer's plant. We assume that this capacity is fixed during the horizon. No upward/downward adjustments are allowed. In our model, capacity level directly impacts the value of spot market trading as shown in the following experiment. The first experiment is conducted when the capacity is low (i.e., $K = 10$ in the base case). Figures 7.2 and 7.3 demonstrate the value of the spot market trading in low/high capacity levels. The interesting managerial insight is as follows: The value of spot market trading is high when the capacity level is low. This value increases as spot market price increases and backlogs are higher. As the spot market price decreases and inventory levels are higher, the value of spot market trading diminishes.

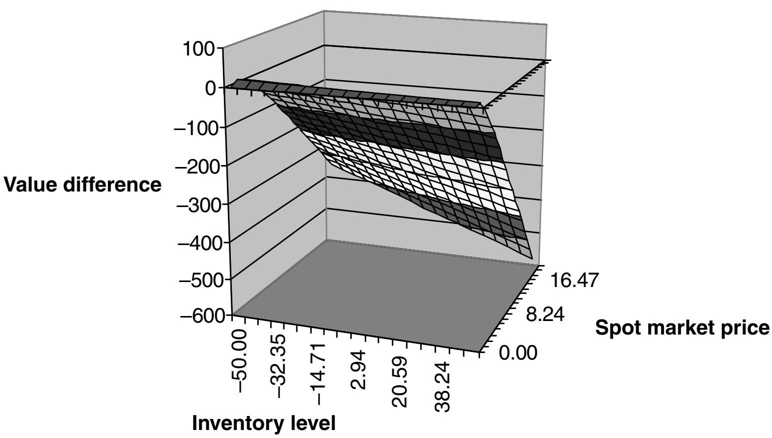


FIGURE 7.3 Value of spot market trading: High (loose) capacity, ($K = 300$).

The second experiment is conducted for high capacity in order to compare and contrast the two cases. It is shown that, as also visible in the Figure 7.3, the value is lower than the low capacity case. In addition, as the spot market price increases, the value drops much faster than in the low capacity case. Only very low spot market prices generate real value. Having ample capacity, the producer is first willing to use her own production facility to produce the commodity before going to the spot market. Nevertheless, very low spot prices may be a good opportunity not to miss for spot market procurement. In summary, production capacity needs to be closely monitored while deciding to trade in the spot market.

7.3.3.3 Impact of Production Cost. In this experiment, we examine the impact of the production cost on the shape of the value function and the optimal control policy. We also delineate under which conditions the spot market trading is valuable. The main insight we derive is as follows: As production becomes more expensive for the producer, more of the commodity is procured from the spot market. The value function becomes more flat as the unit production cost decreases since less trading is done in the spot market as most of the long-term contract demand is fulfilled by production. When the production cost goes to zero in the limit, no purchase is made in the spot market as expected.

The value of spot market trading increases as the production cost goes up since more trading is performed profitably in the market rather than producing in the facilities. When the unit production cost becomes very high, production ceases completely and spot market becomes the sole source of procurement and sales for the producer. Moreover, the producer reacts similarly to production cost increase and capacity decrease. Hence, higher spot prices favor more selling in the market and incurring backordering costs as lower spot prices favor purchasing from the spot market and incurring holding costs. As the production costs depreciate, the value of spot trading is enhanced when the spot market prices are moving to

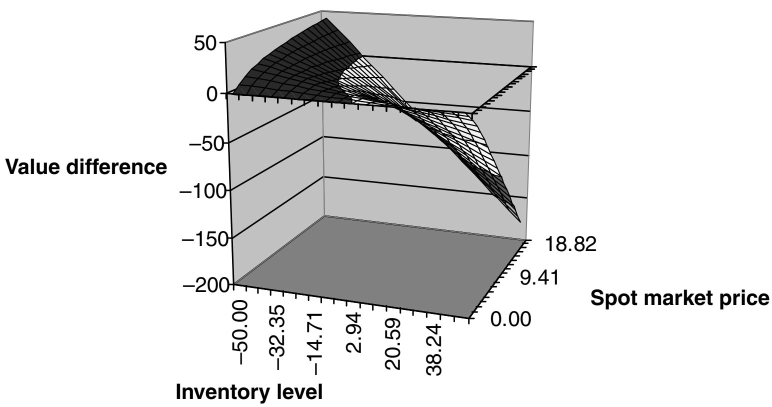


FIGURE 7.4 Value of spot market trading: Low production cost, ($c = 2$).

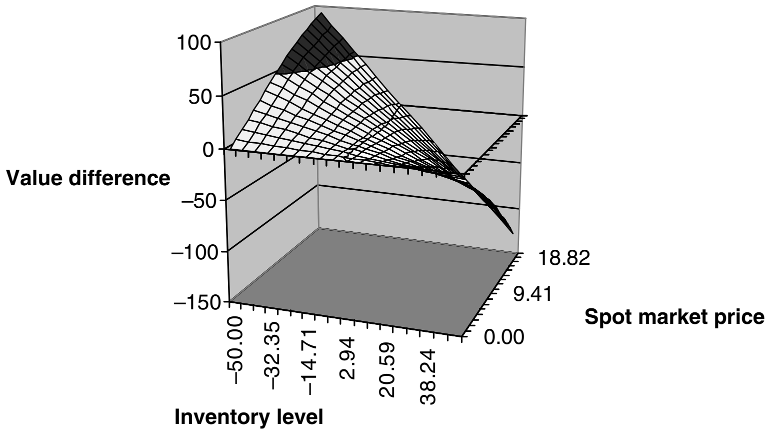


FIGURE 7.5 Value of spot market trading: High production cost, ($c = 10$).

higher levels as more selling in the spot market is expected. Figures 7.4 and 7.5 depict the value of the spot market trading in low/high unit production costs.

7.3.3.4 Impact of Inventory Holding and Backordering Costs.

Storage of commodity is a critical process for the producer. Previous literature addressed supply chain contracting problem for nonstorable and capital-intensive commodities, such as electricity and airline seats. Representative work includes Wu et al. (2002) and Dong and Liu (2007). First, we test for the different values of the holding cost that are given in Table 7.1. Keeping other parameters constant, increasing the inventory holding cost affects neither the value function nor the optimal control policy. We conjecture that the producer does not hold inventory between periods as the optimal policy has bang-bang control structure. Also note that speculation does not yield profit, because the spot price is correctly priced. If it is profitable to sell in the future, it must also be profitable to sell now and get the profit immediately.

Second, the impact of backordering cost is explored. Observe that the backordering process takes place when long-term contract demand cannot be fully satisfied due to low inventory levels at any time in the horizon. Despite the similarity of this cost to the inventory holding cost, the results are different. We test with the following values of backordering costs, $b = 0.5, 2, 5, 10$. First we observe that the overall cost levels go up as the backordering cost increases. Furthermore, the curvature of the value function changes. The value function becomes more linear as the backordering cost increases. Decision for the optimal spot trading and production is similar when the backordering cost is higher than the base value (i.e., $b = 1$). The producer uses all of the production capacity to produce and sell in the spot market.

Next, we conduct two experiments with low, $b = 2$, and high, $b = 10$, unit backordering cost values to delineate the value of spot market trading. We obtained

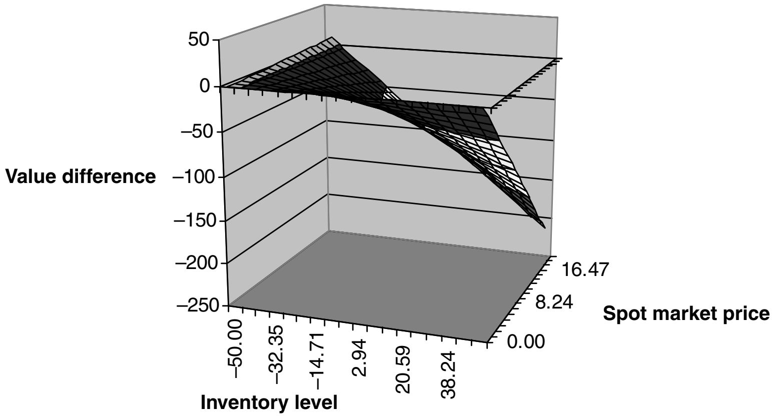


FIGURE 7.6 **Value of spot market trading: Low backordering cost, ($b = 2$).**

the following insights: When fulfilling the backlogs becomes more expensive, the producer becomes more willing to purchase from the spot market. Note that for high backlog levels, the producer is better off trading in the market if the backordering cost is low. However, when backordering cost increases, the producer becomes less willing to sell to the spot market when the spot price is high. As the inventory levels go up, value of spot market trading diminishes with increasing spot market price. In that case, the producer has more incentive to fulfill the long-term contract demand by producing at her own facility. Please note the difference in the values of the spot market trading for low and high backordering cost cases in Figures 7.6 and 7.7.

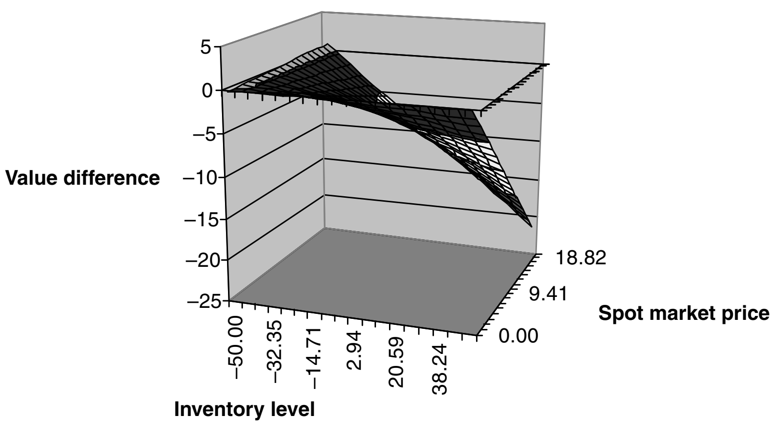


FIGURE 7.7 **Value of spot market trading: High backordering cost, ($b = 10$).**

7.3.3.5 Impact of Risk-Free Interest Rate. We obtain the following results for the impact of risk-free interest rate on the optimal policy and value of spot market trading. As the risk-free interest rate increases, the value of spot market trading increases. The intuition is that the discounting effect of the future time periods becomes more pronounced in case of high risk-free interest rate. Moreover, asymmetric saddle shape of the value function characterizes the nature of the optimal policy as follows: The producer gains more when purchasing from the spot market with low spot price than selling to the spot market with high spot price. This same relationship, which was shown to hold in the previous sections for low capacity and high production costs, is still valid for any risk-free interest rate. However, the impact on the total costs is larger when the risk-free interest is higher (i.e., for higher interest rates), the rule mentioned previously regarding the production capacity yields less costs, thus higher spot market value for the producer.

7.3.3.6 Impact of Spot Price Volatility. We examine the impact of spot price volatility on the value of the spot market trading. In the spirit of real options, if one considers the spot market trading as an option that the producer has the right to exercise at any time, then volatility of the spot market price will have an impact on the optimal exercise policy of this option.

TABLE 7.2 **Summary of Main Insights**

Changes in Key Levers	Impacts
Spot price ↓	Value of additional capacity ↑
Spot price ↑	Favors spot market trading
Additional capacity	Favors purchasing from the spot market
	Track the spot prices carefully as it affects the inventory
Capacity is low	Value of spot market trading ↑
	Value ↑ as spot price ↑ and backlogs are high
	Value ↓ as spot price ↓ and inventory is high
Capacity is high	Value of spot market trading is lower than low capacity case
	Value ↓ as spot price ↑
	Very low spot prices induce purchasing from the market
Risk-free interest rate ↑	Value of spot market trading ↑
Backordering cost ↑	Value of spot market trading ↑
	As backlogs ↑, value of trading ↑
Production cost ↑	Value of spot market trading ↑
	Producer likely to use market for procurement than selling
	When backlogs are highest, value of spot market trading highest
Spot price volatility ↑	Value of spot market trading ↑
	Impact is visible when volatility is very high
Need to hold inventory	Zero as no economic rationale

Our experiments show that the high volatility will increase the value of spot market trading. However, this impact is much smaller than the impacts of the previous key levers such as production capacity or unit production cost. One possible explanation is as follows. All other key levers affect the objective function linearly except the spot price volatility. Spot price volatility's impact is more subtle due to its inherent randomness. We do not clearly know the fluctuations in the spot price before they are realized. It would be more informative to run simulations instead of using fixed volatility values. This is indeed necessary in practice to be able to obtain robust results. Yet, our general intuition based on real options literature continues to hold. Increase in variability for the uncertain factor enhances the option value related to that factor. Table 7.2 provides a summary of the main insights obtained from the numerical study.

7.4 Conclusion

In this chapter, we addressed the integrated production and risk hedging decisions of a risk averse firm in single and multiperiod settings. We demonstrated the value and benefits of risk hedging using forwards/futures and options as well as spot market trading. We also presented a decision model for a commodity producer who can use the spot market trading option besides production. One of the main results of our analysis is that while supplying via long-term contract, trading in the spot market should be considered as an option for the producer. Then, both selling and purchasing in the spot market can be used when beneficial. We showed that the value of trading in the spot market (i.e., the option value of trading) depends on some critical factors: production capacity, production cost, backordering cost, backlog levels, spot price, and its volatility. Based on these factors and how they change over time, decision guidelines can be developed for managers to help them decide whether the trading in the spot market is worthwhile.

In multiperiod models, supply chain managers needs to be careful about the resolution of price uncertainty and timing of operational decisions. We clearly observe that the capital investment and resource allocation decisions need to be made in conjunction with the production decisions for agricultural products. In those industries, managers do not have much production flexibility once the seeds are planted, hence hedging becomes more critical during the later periods until the harvest ends. Furthermore, the yield risk in production may depend on weather, which is not controlled by any firm. Thus, hedging with forwards/futures as well as options including weather derivatives would be beneficial. In contrast, for industries such as fashion retail or consumer electronics, products have shorter life cycles, yet forecasting the demand is much harder. Production volume flexibility—having secondary production options along the way—and timing flexibility—postponing the production—is found to be quite useful operational hedges as more reliable demand information is acquired (see Van-Mieghem 2003 for an excellent review). On top of these operational hedges, hedging with financial instruments would be useful in cases where there are reserved production capacities (probably offshored or outsourced) and not using them incurs high penalties. This may require the production decision be made prior to the uncertainty resolution. In sum, we expect that the firms where resource allocation and production must be made at the same time are better off hedging price, basis, and yield risks using financial instruments such as forwards/futures and options. Firms that have production volume and/or timing flexibility should consider first using operational hedging when cheaper and carefully analyze the solid benefits of financial hedging before adding to their hedging portfolios.

We list a number of interesting future research directions. Our stochastic control model assumes away a few critical considerations seen in practice. One of these considerations is the shipping delay that may occur between the production facilities and the warehouses where the actual sales occur. This problem mathematically poses a challenge. We believe that insights obtained might be different by incorporating the delay into the model.

Second, market incompleteness and its implications needs to be addressed for this specific problem. No transaction costs assumption in the spot market trading is not realistic taking into account the illiquidity of some of the commodity market exchanges. Therefore, friction in the trading side needs to be incorporated into the models and consequences in the solution have to be examined. Recently, Haksoz and Kadam (2009) presented a model for supply portfolio risk assessment in the presence of transaction costs.

Third, none of our analysis in this chapter addressed the coordination between the seller and the buyer. This coordination problem should be addressed for a commodity supply chain where the spot market offers trading to both players, not only one of the them. Recent work by Dong and Liu (2007) contributes in this line of research. We believe that analysis of these problems can enhance our understanding of hedging and managing risk in global supply networks.

REFERENCES

Anderson R.W., Danthine, J.P. 1981. Cross hedging. *Journal of Political Economy*, **89**(6):1182–1196.

Anderson R.W., Danthine, J.P. 1983. The time pattern of hedging and the volatility of futures prices. *Review of Economic Studies* **50**(2):249–266.

Antonovitz F., Nelson, R.D. 1988. Forward and futures markets and the competitive firm under price uncertainty. *Southern Economic Journal*, **55**(1):182–195.

Babich V., Burnetas A.N., Ritchken P.H. 2007. Competition and diversification effects in supply chains with supplier default risk. *Manufacturing & Service Operations Management*, **9**(2):123–146.

Batra R.N., Ullah A. 1974. Competitive firm and the theory of input demand under price uncertainty. *Journal of Political Economy*, **82**(3):537–548.

Bowden R.J. 1995. Production organisation and risk control when market instruments are available. *Management Science*, **41**(6):1073–1082.

Brennan M.J., Schwartz, E.S. 1985, Evaluating Natural Resources Investments, *Journal of Business*, **58**:135–157.

Caldentey R., Haugh M. 2006. Optimal control and hedging of operations in the presence of financial markets. *Mathematics of Operations Research*, **31**(2):285–304.

Chen X., Sim M., Simchi-Levi D., Sun P. 2007. Risk aversion in inventory management. *Operations Research*, **55**(5):828–842.

Chod J., Rudi N., Van-Mieghem J.A. 2010. Operational Flexibility and Financial Hedging: Complements or Substitutes?. *Management Science*, **56**(6):1030–1045.

Dada M., Petruzzi N.C., Schwarz, L.B. 2007. A Newsvendor's Procurement Problem when Suppliers Are Unreliable, *Manufacturing & Service Operations Management* **9**(1),9–32.

Ding Q., Dong L., Kouvelis P. 2007. On the integration of production and financial hedging decisions in global markets. *Operations Research*, **55**(3):470–489.

Dong L., Liu H. 2007. Equilibrium forward contracts on nonstorable commodities in the presence of market power. *Operations Research*, **55**:128–145.

Eberly J.C., Van-Mieghem J.A. 1997. Multi-factor dynamic investment under uncertainty. *Journal of Economic Theory*, **75**:345–387.

Feder G., Just R.E., Schmitz A.: 1980. Futures Markets and the Theory of the Firm Under Price Uncertainty, *Quarterly Journal of Economics*, **94**(2):317–328.

Frechette D.L. 2001. The Demand for hedging with futures and options. *Journal of Futures Markets*, **21**(8):693–712.

Gaur V. and Seshadri S. 2005. Hedging inventory risk through market instruments, *Manufacturing & Service Operations Management*, **7**(2):103–120.

Global Risks 2009: A Global Risk Network Report: 2009. Retrieved from http://www.weforum.org/pdf/globalrisk/globalrisks09/.

Gollier C. 2001. *The Economics of Risk and Time*, The MIT Press, Cambridge, Massachusetts.

Hadjidimos A., Houstis E.N., Rice J.R., Vavalis E. 1999. Analysis of iterative line spline collocation methods for elliptic partial differential equations. *SIAM Journal on Matrix Analysis and Applications*, **21**:508–521.

Haksöz, Ç, Kadam, A. 2008. Supply risk in fragile contracts. *MIT Sloan Management Review*, **49**(2):7–8.

Haksöz Ç, Kadam, A. 2009. Supply Portfolio Risk. *The Journal of Operational Risk*, **4**(1):59–77.

Haksöz Ç, Seshadri, S. 2007. Supply chain operations in the presence of a spot market: A review with discussion. *Journal of Operational Research Society*, **58**(11):1412–1429.

Holthausen D.M. 1979. Hedging and the competitive firm under price uncertainty. *American Economic Review*, **69**(5):989–995.

Hull J.C. 2003. *Options, Futures, and Other Derivatives*. Englewood Cliffs, New Jersey: Prentice-Hall.

Johnson L.L. 1960, The Theory of Hedging and Speculation in Commodity Futures. *The Review of Economic Studies*, **27**:139–151.

Kamrad B., Ritchken P. 1991. Multinomial approximating models for options with k state variables. *Management Science*, **37**:1640–1652.

Kushner H.J. 1977. *Probability Methods for Approximations in Stochastic Control and for Elliptic Equations*. New York: Academic Press.

Kushner H.J., Dupuis, P. 2001. *Numerical Methods for Stochastic Control Problems in Continuous Time*. New York: Springer-Verlag.

Lapan H., Moschini G., Hanson S.D. 1991. Production, hedging, and speculative decisions with options and futures markets. *American Journal of Agricultural Economics*, **73**(1):66–74.

Lence S.H. 1995. The economic value of minimum variance hedges. *American Journal of Agricultural Economics*, **73**(1):66–74.

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

Matthews R.G. 2008a, Corporate news: Steelmakers squeeze suppliers—as weak demand weighs on raw-material prices, mills look to jettison contracts. *The Wall Street Journal*, November 18.

Matthews R.G. 2008b, Steel surcharges draw ire. *The Wall Street Journal Europe*, July 7.

Miranda M.J., Fackler P.L. 2002. *Applied Computational Economics and Finance*, Cambridge, Massachusetts: The MIT Press.

Moschini G., Lapan H. 1992. Hedging price risk with options and futures for the competitive firm with production flexibility. *International Economic Review*, **33**(3): 607–618.

Moschini G., Lapan H. 1995. The hedging role of options and futures under joint price, basis, and production risk. *International Economic Review*, **36**(4):1025–1049.

Oksendal B. 2005. *Stochastic Differential Equations: An Introduction with Applications*. Berlin Heidelberg, Germany: Springer-Verlag.

Rothschild M., Stiglitz J. 1970. Increasing risk I: A definition. *Journal of Economic Theory*, **2**(3):22–243.

Sandmo A. 1971. On the theory of the competitive firm under price uncertainty. *American Economic Review*, **61**:65–73.

Schwartz E.S. 1997. The Stochastic behavior of commodity prices: Implications for valuation and hedging. *Journal of Finance*, **52**:923–973.

Sethi S., Thompson G.L. 2000. *Optimal Control Theory: Applications to Management Science*. Norwell, Massachusetts: Kluwer Academic Press.

Sezer A.D., Haksöz Ç. 2010. Optimal decision rules for product recalls. Working paper. Istanbul, Turkey: Sabancı University, Faculty of Management.

Sheffi Y. 2007. *The Resilient Enterprise: Overcoming Vulnerability for Competitive Advantage*, The MIT Press, Boston, M.

Stein J.L. 1961. The Simultaneous determination of spot and futures prices. *American Economic Review*, **51**(5):1012–1025.

Tang C.S. 2006. Perspectives in supply chain risk management. *International Journal of Production Economics*, **103**:451–488.

Tang C.S. 2008. Making products safe: Process and challenges. *International Commerce Review*, **8**(1):49–55.

Tevelson R.M., Ross, J., Paranikas P. 2007. A key to smart sourcing: Hedging against risk. *The Boston Consulting Group publication*?.

Tomlin, B. 2009, Impact of supply learning when suppliers are unreliable. *manufacturing & Service Operations Management*, **11**(2):192–209.

Tsompanopoulou, P., Vavalis, E. 1998. ADI methods for cubic spline collocation discretizations of elliptic PDEs, *SIAM Journal of Scientific Computing*, **19**:341–363.

Van-Mieghem J.A. 2003. Capacity management, investment and hedging: Review and recent developments, *Manufacturing & Service Operations Management*, **5**:269–302.

Van-Mieghem J.A. 2007. Risk mitigation in newsvendor networks: Resource diversification, flexibility, sharing, and hedging, *Management Science*, **53**(8),1269–1288.

Working H. 1953. Hedging reconsidered, *Journal of Farm Economics* **35**(4):544–561.

Wu D.J., Kleindorfer P., Zhang J.E. 2002. Optimal bidding and contracting strategies for capital-igoods, *European Journal of Operational Research*, **137**:657–676.

CHAPTER EIGHT

Capacity Expansion As A Contingent Claim: Flexibility And Real Options In Operations

BARDIA KAMRAD

Contingent claims methodology have had a considerable impact in the integration of financial risk management concepts in operations and related capital allocation problems. The methodology has also helped in expanding the set of needed tools for evaluating risk-based decisions through providing an alternative economic platform for analyses. This chapter provides a prefatory introduction to these concepts by reviewing their application in the context of a capacity investment problem where the flexibility (real option) to expand capacity remains a future alternative. The strategic insights and the intuition resulting from the approach are detailed and calibrated relative to the more established decision analysis approach.

Handbook of Integrated Risk Management in Global Supply Chains, First Edition. Edited by Panos Kouvelis, Lingxiu Dong, Onur Boyabatli, and Rong Li.

8.1 Introduction

Discounted cash flow (DCF) methods are the most prevalent valuation framework for analyzing capital investment opportunities. Several reasons justify this popularity. The methodology is economically sound and mathematically straightforward, easily implementable, and it offers clear-cut decision rules. Investments are economically acceptable if conventional hurdles of positive net present value, (NPV) is greater than 0 or internal rate of return (IRR) exceeding cost of capital are met. DCF methodology also provides an alternative approach to the more established payback and accounting rate of return methods.[1] As in most capital budgeting techniques, however, the accuracy of the resulting NPV (IRR) is directly dependent on the accuracy of the cash flow forecasts and the risk adjusted discount rate estimates (Brealy and Myers 2000, Copeland and Antikarov 2001).

More exclusively, however, DCF methods are prone to innate structural deficiencies in presence of uncertainty, operating flexibility, and irreversibility. The latter is in reference to the nonrecoverable (or partially recoverable) investment outlays; essentially treated as sunk costs in that they have negligible salvage value. These characteristics significantly impact the risk structure within and across investment projects, have value altering and strategic implications, and change the economic interpretation and insight to the analyses. Furthermore, their omission may also induce a systematic undervaluation problem that can influence the desirability, the level, and the timing of investments considered. Yet, these omissions reflect the deficiency of a particular valuation methodology employed and not their quantitative underpinnings.

Ideally, the requisite approach would entail a valuation mechanism endogenizing flexibility, volatility, contingency, and irreversibility into a simple numerical result (e.g., NPV) and a clear-cut decision outcome (e.g., NPV > 0). In this vein, real options valuation (ROV) extends existing practices in a number of distinct albeit related, ways.[2] First, and similar to DCF, the ROV approach results in an expected net present value with the exception that uncertainty is explicitly incorporated into the valuation process. Second, operating flexibility is directly assimilated into the analysis and the additional value (*flexibility premium*) derived from an existing set of operating options (managerial control) is quantified. Third, when used under "appropriate or intended" conditions, the approach can differentiate between upside potential and downside risk. The term *appropriate* is used in reference to the set of requisite market-related conditions, reviewed shortly, to correctly implement the approach. Thus, the existence of operating options within an investment project provides the necessary means to adapt to uncertainty such that the upside potential is improved while downside losses are limited. Viewed in this light, large volatility, in some cases, may be beneficial as

[1] See Ryan and Ryan (2002).

[2] Stewart Myers (MIT) coined the term "real options" in the late 1970s. It provides a contrast between financial and real options the same way we distinguish between financial and *real* assets.

it expands the upside potential without detriment to the downside. Fourth, the ROV approach can entrench the hierarchical and sequential nature of investment decisions under uncertainty. These characteristics have nontrivial implications. The strategic options associated with an investment opportunity typically result from interdependencies with future sequential investments. They may be viewed, in the words of Myers (1984), as "cross-sectional" and "time-series" links between projects, respectively. That is, the current investment opportunities' impact on the firm's cash flows and assets versus the impact on the future follow-up opportunities. In this context, by not committing to a current opportunity (e.g., a negative NPV project), other future opportunities may be lost or, at least delayed.

Deliberating these interdependencies has some decision makers advocating lower discount rates to avoid penalizing investments with implicit strategic options. Others find it justifiable to accept negative NPV projects simply on the basis of their potential to create subsequent follow up and growth opportunities. Surely, the value of retaining options to other future investments (growth options) is subtly embedded in the value of strategic options. Precisely for these reasons, DCF methods are unable to value sequential investments. The methodology is not structured to value options.

This concept is particularly important in operational environments where investments are typically lumpy, sequential, and contingent on earlier investments (e.g., capacity expansion; new plant construction; supply contracts, product architecture, product life cycle analysis, modularity and upgrades, global manufacturing networks, etc). We will address this issue more explicitly in the context of an upcoming numerical example. Indisputably, decision tree (analysis DTA) and simulation methods can be employed to incorporate uncertainty, to account for flexibility and the strategic implications of sequential investments. However, the basic foundation to these techniques rests on the economic principles of decision analysis and discounted cash flow methods.

DTA is an instinctive, powerful, and a relatively straightforward approach for the purposes of integrating uncertainty and flexibility in the valuation process. With classical decision tree techniques, typically the utility function of the decision maker, reflecting preferences or aversions toward reward and risk, enters the valuation process. Also, future scenarios with corresponding (subjective) probabilities and cash flow estimates are required. In application terms, however, as the number of potential future scenarios increases, so will the complexity in specifying scenario-dependent probabilities. This concern also includes scenario-contingent cash flow forecasts. A further concern is the implied "equi-risk" treatment of scenario-contingent cash flows or payoffs. In particular, by assuming a constant risk adjusted discount rate while working back through the tree, scenario-specific payoffs are tacitly assumed to be in the same financial risk class.

Simulation methods are also an irrefutably powerful evaluative tool. Their application in the current context has also increased in recent years, in no small measure due to advances in computing capabilities. A potential concern in their use, however, is a simple matter of transformation(i.e., mapping simulation outputs into meaningful managerial actions.) Specifically, transforming the probability distribution of the resulting NPVs into tactical and strategic actions is a nontrivial undertaking. A more practical alternative might be one of simulating the cash flows associated with the investment project in question. The resulting probability distribution can then be used to estimate the variance of the cash flows. Using this variance estimate, an appropriate risk-adjusted discount rate can be derived to establish an expected NPV for the project. This alternative approach offers an unambiguous criterion for the investment decision (see also Brealy and Myers 2000).

The preceding review is not a broad brush criticism of the techniques mentioned. Based on the set of assumptions that characterize its fundamentals each method offers unique advantages in its aptness and involves specific drawbacks. Nor the intent is to imply that conventional DCF methods routinely undervalue all investment projects. In fact, for investments where the scope of managerial actions tends to be limited, and where forecasted cash flows suggest negligible volatility (similar to a bond), the DCF approach is the right means to analysis. Nor we are implying that DTA and simulation methods are so fraught by complications that their use for analysis is more problematic than practical. We are also not suggesting that the ROV approach is the ideal means to analysis of investment decisions in the face of uncertainty. As in other techniques, the ROV methodology has its advantages and its share of drawbacks, assumption-based impracticalities, application restrictions, and implementation complexities and limitations. This will be established in the context of an upcoming example, shortly. Nonetheless, the ROV approach offers theoretical and practical advantages that make its adoption an insightful choice in many situations. It is worthwhile to note that in any dynamic valuation environment, no single approach will be optimal. The appropriate choice, and a recommended approach, is to employ all pertinent valuation methods to address the decision problem. This provides a basis to compare the resulting method-specific NPVs and among other considerations, to assess the reasons for their discrepancies.

In the following sections, we proceed by reviewing the Black and Scholes (1973) and Merton (1973) option pricing model. Options are a particular form of contingent claims. Generally, an asset whose value is contingent on the random prices of one or more security is defined as a contingent claim. For our purposes, the case of a plain vanilla European call option is demonstrated here without the usual financial and mathematical trimmings that precisely illustrate the model's derivation.[3] This is done in the interest of simplicity and brevity. The suggested references including Hull (2008), and Jarrow and Rudd (1983) are recommended as further readings in providing a more complete treatment of the approach.

[3] The holder (owner) of a European *call* option has the right to *purchase* an underlying asset (typically a stock) at a specific time (maturity or exercise date) and a pre-established price. The owner of a European *put* option has the right to *sell* an underlying asset (typically a stock) at a specific time (maturity or exercise date) and a pre-established price. If an option (put or call) is exercisable at any time during its maturity interval it is referred to as an American option.

8.2 A Financial Option Pricing Model: Black Scholes (1973) and Merton (1973) Model

The objective in this section is to provide a prefatory understanding of the fundamental features of this well-known model. The model's understanding is essential in demonstrating how the ROV approach works. To this end, it should be noted that the Black and Scholes (1973) and Merton (1973) option pricing model (henceforth, B–S & M [1973]) can be derived in an alternative way than presented here.[4] Fisher Black and Myron Scholes (1973) and Robert Merton (1973) independently arrived at the same pricing formula governing the value of an option. The assumptions required to derive the B–S & M (1973) model are:

1. Risk free rate of return, r is constant during the option's maturity [0, T]: T is the maturity date.
2. Markets for stocks, bonds and options are frictionless (i.e., no transaction costs, no restrictions on short sales [selling borrowed assets], no taxes, and shares of the stock are infinitely divisible).
3. Stock price process, $S(t)$ is given by a geometric Wiener process (GWP) over the option's maturity interval. That is,

$$dS(t) = (\alpha - \delta)S(t)dt + \sigma S(t)dW(t) \tag{8.1}$$

The given input parameters to the model α, δ, and σ; are all assumed constant over the maturity interval $[0,\ T]$. The total expected return rate (capital gains and dividends) is given by the constant, α; δ is the constant dividend yield and σ is the constant standard deviation of the price return.[5] In equation ($1), $dW(t)$ is

[4] It is also possible to derive the B–S & M (1973) through constructing a (self-financing replicating) portfolio consisting of a short position (i.e., borrowing) in T-bills and a long position in a *number* of shares of the stock (i.e., the option delta, as defined later). The portfolio is constructed in such way that all the cash inflow/outflow from this portfolio perfectly duplicates the option's cash flows during the maturity interval, after the initial construction of the portfolio, there is no need to make any cash inflow/outflow adjustments (self-financing). Note also that the portfolio has the same exposure to the stock as does the option to be valued. Therefore, to avoid arbitrage opportunities, the option and the portfolio must command the same value. We say that the market is "complete" when a contingent claim's value can be perfectly replicated through a combination of existing traded assets in the market at every point in time and for all possible states of nature. If the market is complete for a contingent claim of interest, it is always possible to construct a replicating portfolio, option is redundant. However, options create tremendous economic value in terms of trading and financial market activity.

[5] Stocks do not pay dividends in yield form, the use of dividend yield is a mathematical convenience. For an American option, if the dollar dividend paid does not justify an early exercise, then the B–S & M (1973) can be employed, effectively valuing a "European" contract. However, if the dollar dividend justifies the early exercise of the option, the B–S & M (1973) cannot be used to value the option.

an increment in the standard Wiener process, $W(t)$. [6] The GWP assumption for the price return process also implies that conditional on the time zero stock price $S(0)$, the stock price, $S(t)$, is log-normally distributed:

$$S(t) = S(0)e^{\left(\alpha - \delta - \frac{\sigma^2}{2}\right)(t) + \sigma W(t)}$$

Thus, negative prices are precluded (with probability one), which is a desirable property for a stock price model. Note that, $W(t) \sim \mathbb{N}(0, t)$ implying that $W(t)$is normally distributed with mean zero and variance t. In essence, this assumption implies that on an average basis the stock price increases exponentially over time with random perturbations around this average growth.

To this end, let $C(S(t), t) \equiv C(S, t)$ define the call value at time $t$as a function of the underlying asset's value and time. The function $C(S(t), t)$ is assumed to be twice differentiable in $S(t)$ and once differentiable in t. The instantaneous change in the call's value over an instant of time, dt, results in a second order partial differential equation (p.d.e.) that is to be solved subject to the call option's terminal condition. The solution to this p.d.e. is the B–S & M (1973) valuation model. In particular, for $t \in [0, T]$ and $0 \leq S(t) \leq \infty$:

$$\frac{\partial C(S, t)}{\partial S(t)} S(t)(r - \delta) + \frac{1}{2}\frac{\partial^2 C(S, t)}{\partial S^2(t)} S^2(t)\sigma^2 + \frac{\partial C(S, t)}{\partial(t)} = rC(S, t) \qquad (8.2)$$

$$C(S(T), T) = \text{Max}(S(T) - X, 0) \qquad (8.3)$$

This solution is:

$$C(S, 0) = C_0 = S(0)e^{-\delta T}\Phi(d_1) - Xe^{-rT}\Phi(d_2) \qquad (8.4)$$

with,

$$d_1 = \frac{\ln\left(\frac{S(0)}{X}\right) + \left(r - \delta + \frac{\sigma^2}{2}\right)T}{\sigma\sqrt{T}}; \quad d_2 = \frac{\ln\left(\frac{S(0)}{X}\right) + \left(r - \delta - \frac{\sigma^2}{2}\right)T}{\sigma\sqrt{T}}$$
$$= d_1 - \sigma\sqrt{T} \qquad (8.5)$$

In the above equations, C_0 is the B–S & M (1973) European call option value at time zero; $\Phi(d)$ is the cumulative standard normal probability (i.e., $P[Z \leq d]$ where $Z \sim \mathbb{N}(0, 1)$). Given the above nomenclature, we are now in a position to provide the needed intuition to this model. This is done in the next section.

[6] We have: (1) $dW(t) \sim N \quad 0, dt$; (2) $E \quad dW(t) = 0$; (3) $Vard \quad W(t) = dt$; and (4) $Cov \quad dW(s), dW(t) = 0,\ s < t$, That is, serial independence.

8.2.1 MODEL INSIGHT AND SOME INTUITION

1. Equation (8.2) is commonly referred to as the *fundamental* equation. An unusual feature is the second order effect: proportional to the volatility of the stock.[7]

2. Equation (8.4) is the solution to equation (8.2), subject to the terminal condition which is given by equation (8.3) and the non-negativity constraint for the stock price. In reference to equation (8.4), it is important to note the following:

- The total expected return rate of the stock does not appear in the model. That is, the option's value is independent of, α. Note too, that the option valuation process is also independent of exogenously provided preference or utility functions. It is commonly recognized that investors require a higher expected return reflecting a premium as compensation for higher levels of risk.[8] However, in the current context, preferences and aversions toward risk actually do not enter the valuation process. This "risk-preference" independence feature of the B–S & M (1973) is important and as will be discussed shortly, lends itself to a well-established concept known as the risk neutral valuation principle formalized by Cox and Rubinstein (1976). Consistent with footnote 4, if the market is complete (i.e., a replicating portfolio can be constructed), then risk neutral valuation becomes possible. In Section 8.2.2, a modestly detailed review of the risk neutral valuation concept is provided.
- The B–S & M (1973) model (i.e., equation (8.4)) requires five input parameters assumed to be *given*. These parameters are: (1) initial stock price, $S(0)$; (2) exercise price, X; (3) maturity time, T; (4) risk free rate of return, r; and (5) volatility (standard deviation of the stock price return), σ. Note that to obtain results no information or input regarding preferences or aversions toward risk (i.e., utility functions) is required. As stated earlier, α does not affect the option's value. With the exception of the volatility parameter, all other inputs are readily available in financial publications such as the WSJ.
- In equation (8.4), the dividend yield is taken out of (in effect discounted from) the stock price. Since any dividend paid out accrues to the owner of the stock and not the owner of the option. This adjustment is necessary to avoid arbitrage opportunities (riskless profits can made).
- $\Phi(d_2)$ is the probability that at its maturity the call option is exercised. That is, $\Phi(d_2) = P\{S(T) > X\}$. In other words, the probability of ending up in the money. Note that, $\Phi(d_2) = P(Z < d_2)$ with $Z \sim \mathbb{N}(0, 1)$.

[7] The partials with respect to the stock price and time are to be expected. The second order effect results from "Ito calculus" or stochastic calculus. To be exact, we require function $C(S(t), t)$ to be Ito differentiable in that the function admits the second order effect.

[8] In the derivation of the *fundamental* equation, p.d.e. (8.2), the term α is eliminated.

- $\Phi(d_1)$ is also a probability but does not enjoy the crisp interpretation of the above-mentioned exercise probability. It is commonly referred to as the *perfect hedge ratio* or the *call option delta*: $\Phi(d_1) = \frac{\partial C}{\partial S} = \Delta$
  - The *perfect hedge ratio*: A long holding of $\Phi(d_1) = \frac{\partial C}{\partial S}$ shares of stock perfectly offsets a written call. In other words, the call's exposure to the underlying stock. The perfect hedge ratio is riskless and the funds invested in it earn the riskless rate.
  - $\Phi(d_1)$ also indicates the call's sensitivity to the stock price: A \$1.00 change in the price of the stock will lead to a $\$\Phi(d_1)$ in the call value. Note that for most situations, $\Phi(d_1) < 1.0$.[9]

3. Dividend yield is the dollar dividend divided by the stock price. Stocks pay dollar dividends, not dividend yield. In this sense, a shortcoming of the B–S & M (1973) model is the model's inability to capture the early exercise feature of "American" style options when dollar dividends are paid out and early exercise of the option is economically justifiable. In these situations, the closed form analytic solutions obtained for "European" options do not exist and typically numerical methods are employed to obtain results.
4. An interpretation of the above equation (8.4) is that the call option value is effectively an expected net present value, conditional on the call expiring in the money at maturity (i.e., $S(T) > X$). In other words, the call value represents the present value of the expected stock price at maturity (given that it exceeds the exercise price), less the present value of the expected cost of exercising the option.[10] Another interpretation, one that corresponds to the perfect hedge ratio (or the option delta) explanation, is as follows. The call value represents the dollar purchase of the stock (long position) less the borrowed amount (short position). In this case, the long position in the stock is in reference to the long holding of $\Phi(d_1)$ shares multiplied by the current stock price.
5. With modest alterations, a European put option value can be obtained. Without loss of generality, we have the put's terminal condition as:

$$P(S(T), T) = Max(X - S(T), 0) \tag{8.6}$$

$$P(S, 0) = P_0 = -S(0)e^{-\delta T}\Phi(-d_1) + Xe^{-rT}\Phi(-d_2) \tag{8.7}$$

Both d_1 and d_2 are given by equation (8.5). A word of caution may be in order: In this particular case, the perfect hedge ratio (or the put option delta) is $-\Phi(-d_1) < 0$. Nonetheless, $\Phi(-d_2)$ remains as before in its interpretation: It defines the exercise probability.

---

[9] The condition is true so long as $C(S, t)$ is convex in $S(t)$. See Hull (2008); also see Jarrow and Rudd (1983).

[10] Technically, $PV[E(S(T)|S(T) > X) .P(S(T) > X)] = S(0)e^{-\delta T}\Phi(d_1)$. PV(.) is the present value operator.

8.2.2 RISK NEUTRALITY

Underpinning the concept of risk neutral valuation is the ability to construct a perfect hedge. Essentially, this concept implies that if a perfect hedge can be constructed, then the option's price can be established independent of the underlying stock's expected return. Note that this was the case with equations (8.2), (8.4), and (8.5). That is, the expected return on the stock α does not appear in the fundamental valuation equation or its corresponding call option price solution. In general, in valuing any contingent claim, so long as a perfect hedge can be constructed, risk neutral valuation becomes feasible. That is, the value of the claim will be independent of the underlying asset's expected return.

It is commonly acknowledged that investors are risk averse, thus requiring a premium as compensation for the risk they tolerate. Aversion toward risk by investors implies different expected returns across securities. The absence of the stock's expected return α suggests, however, that the option's value is the same to a risk averse as well as a risk neutral investors. Given this observation, the *risk neutrality argument* advocates option valuation under the simplest of the two preferences: Namely, a risk neutral setting where preferences toward risk are extraneous (Jarrow and Rudd 1993). The natural question then is, when can a perfect hedge be constructed? If the market is "complete" a perfect hedge can always be constructed. The market is "complete" when the option's value can be perfectly replicated through a combination of *existing market traded assets* (stock and Treasury bond), at every point in time, and for all possible states of nature. In general, if the market is complete for a contingent claim of interest (i.e., an underlying traded security and a riskless bond exist, see also footnote 4), it is always possible to construct a replicating portfolio. Hull (2008) and Jarrow and Rudd (1993) provide further readings on this topic.

An intricate, though subtle, implication of risk neutrality argument is that when a perfect hedge is feasible (i.e., the market is complete) then for *valuation purposes only* the total expected return on the underlying stock can be taken as the riskless rate. We caution this is *not to imply* that stock returns (growth rates) are the riskless rate; we know this is not the case. Instead, risk neutrality suggests that when the market is complete, option values can be determined as if the underlying stock returns grow at the riskless rate. For the reader the following two circumstances should remain clear: (1) the implication of risk neutrality for valuation purposes, that is, only for valuation purposes the growth rate in the underlying asset can be taken as if it were the riskless rate; (2) the actual growth rate in the underlying asset can be different (often is) from the riskless rate.

8.3 Real Options Valuation (ROV) in Operations

To most familiar with the term "real options", the phrase and its basic conceptual framework are inevitably linked to the B–S & M (1973) option pricing formula. This is an accurate association. However, the analogy is also restrictive in that it cannot realistically encompass the intricate nature of most operational problems in terms of the complexities involved. Recall that, in effect, the B–S & M (1973) model is fully parameterized in terms of five inputs in order to arrive at a value and all assumed to be given at the outset.

ROV applications to operational problems are wide ranging and for the most part differ substantially in form, complexity, and structure from the plain vanilla option valuation model of the B–S & M (1973). We reference a few by stressing the following papers as essential in gaining familiarity and economic insight: Brennan and Schwatrz (1985), McDonald and Siegel (1985), Pindyck (1991), and for a deeper and more rigorous treatment, Dixit and Pindyck (1994) is suggested.

More operationally focused applications include but are not limited to contractual valuation issues and problems in the analysis of supply contracts: For instance see, Kamrad and Siddique (2004), Li and Kouvelis (1999), and Kamrad and Ritchken (1994). For capacity planning and capacity investment problem applications see He and Pindyck (1993), Birge (2000), Spinler and Huchzermeier (2006), and Kouvelis and Tian (2008). For new product development and product life cycle analysis, Bollen (1999) is suggested. On issues relating to global manufacturing and production networks, production strategy, and production facility ownership structure see, Kogut (1991), Kogut and Kulatilaka (1994), Kulatilaka and Perotti (1998), Huchzermeier and Cohen (1996), Kouvelis, Axarloglou, and Sinha (2001). For production problems with demand as the major source of uncertainty see Chung (1990) and Burnetas and Ritchken (2005), and for applications with price, demand, and yield uncertainty, see Kamrad and Ernst (2001) and Kamrad and Ord (2007). For warranty valuation in light of "system failure" risk and market uncertainty Kamrad and Lele (1998) is suggested. For product mix and product flexibility valuation problems see Anderou (1990) and Hodder and Triantis (1993). Inventory risk management through market-based securities is considered by Gaur and Seshadri (2005). Most applications of real options to operational issues typically entail an optimization problem involving a control variable of interest (e.g., production rate; maintenance rate, etc). As in most optimization problems, various constraints characterizing requirements, limitations, or properties are also par for the course. Some problems may involve more than one state variable (e.g., input price and output demand uncertainty). Others may involve multiple state and control variables. In this sense, as the number of input parameters in a combined "optimization-contingent claims" framework increases so does the problems' complexity. Yet, these modeling complexities cannot be addressed through the standard B–S & M (1973) equation to provide closed form analytical solutions. In most situations, efficient numerical techniques typically defined in terms of a stochastic dynamic program are adapted to provide results.

Notwithstanding these concerns, application of the B–S & M (1973) model in a "real" setting is interesting and insightful. It also provides a deeper understanding of the ROV's potential as a valuation tool in terms of both advantages and disadvantages inherent in the methodology. This approach is demonstrated below through a stylized example.[11] To put matters in perspective, the merits

[11] An adaptation from Timothy Luehrman (1998: HBR-98404): A recommended reading.

along with advantages and disadvantages of this approach are discussed next in reasonable detail.

8.3.1 ILLUSTRATION: ROV, B–S & M (1973) CONTEXT

TIMS Inc., an independent manufacturer of hard drives, is contemplating construction of a production facility. Table 8.1 provides projected revenues and costs, where all information relevant to the analysis including the terminal values is assumed as given. The project requires an initial outlay of \$330.0 million with a phased *capacity expansion alternative* in the fifth year at a cost of \$850 million (\$700 cap ex + \$150 net work cap). It's worthwhile noting that these expenditures (in reference to the expansion costs) are typically discretionary. They reflect the firm's financial, operational, and strategic flexibility in response to market conditions and competition. In this context, it is also important to discern between routine and extraordinary expenditures. The risk adjusted discount rate is assumed 12% per annum. Note that Table 8.1 essentially defines the concept of free cash flows for our purposes. The resulting free cash flows for the immediate eight year horizon, together with the terminal value and assets in year eight are shown in reference lines (8.10) and (8.11), respectively. In effect, a \$45 million per year perpetuity with an assumed constant growth rate of 4% per annum. Using a DCF approach, $NPV_{\text{DCF}} = \$-7.93$ million: implying that under conventional rules the investment opportunity should be rejected.

An implication of the above DCF approach is that regardless of the demand outlook or market conditions, capacity expansion takes place. That is, the option to expand capacity in year five at a cost of \$850 million is exercised with *probability one*, irrespective of its economic viability (see Table 8.1). This implication is not unique to our example. The DCF methodology is not structured to value options: The flexibility premium derived from responding to uncertainty is ignored by the valuation process. In other words, DCF is characteristically static.

Yet, a more supple approach to the investment decision problem might entail a capacity expansion *probability* that its makeup is consistent with the economic lure (or lack thereof) of the expansion phase. For instance, an expansion *probability* signifying the likelihood that the (time-value adjusted) earnings from the expansion phase exceed corresponding expansion costs while taking into account the variability associated with the forecasted earnings.[12] Intuitively, by ignoring cash flow volatility, the unyielding pattern of the cash flows always suggests expansion. To this end, suppose that the investment project is decoupled into two stages. The first stage is the initial capacity installation and construction; the second stage is the follow up capacity expansion phase. We will define these as "Stage 1" and "Stage 2" assets, as a point of reference, respectively.

Note that the latter stage is discretionary and contingent on the former stage. Thus, the initial construction phase is a strategic investment in that it provides

[12] Decision trees can be useful in this regard. The exercise probability can be estimated only if information regarding previous projects with similar characteristics is available. Otherwise, the specification will be subjective.

TABLE 8.1 Cash Flow Projections and DCF Based Net Present Value (the information shown is given)

Year	0	1	2	3	4	5	6	7	8
OPERATING PROJECTIONS:									
1) Revenues:		1,135.05	1,374.75	1,996.56	2,695.92	2,982.15	3,131.61	3,369.90	3,623.70
2) − Cost of goods sold:		845.25	1,027.88	1,486.80	2,010.40	2,223.76	2,338.00	2,509.50	2,698.50
3) = Gross Profit:		289.80	346.87	509.76	685.52	758.39	793.61	860.40	925.20
4) − S G & A:		273.42	322.00	546.00	686.70	723.80	729.96	786.38	849.75
5) = Operating profit:		16.38	24.87	(36.24)	(1.18)	34.59	63.65	74.02	75.45
CASH FLOW CALCULATIONS:									
6) EBIT*(1-tax rate)		10.65	16.17	(23.56)	(0.77)	22.48	41.37	48.11	49.04
7) + Depreciation		47.04	52.01	52.08	126.42	127.54	131.60	133.14	135.38
8) − Capital expenditures	**330.00**	17.22	18.34	19.60	20.58	**700.00**	23.80	24.50	25.48
9) − Increase in net work Cap.	51.80	7.98	8.68	9.66	9.94	**150.00**	11.34	11.48	11.90
10) = Free cash flow	(381.80)	32.49	41.16	(0.74)	95.13	(699.98)	137.83	145.27	147.04
11) + Terminal value, assets*									1,125.00*
Discount Factor by year @12%	1.000	0.893	0.797	0.712	0.636	0.567	0.507	0.452	0.404
12) PV by Year	(381.80)	29.01	32.81	(0.52)	60.46	(397.19)	69.83	65.71	513.76
NPV:	**$ − 7.93 million**								

NOTES:
1. Figures are in Millions of Dollars. Tax Rate is 35%.
2. Risk adjusted discount rate-RADR = 12% p.a. is given.
3. Projected terminal value in the eighth year reflects the present value of perpetuity of $45.00 per year.
* This PV is computed assuming a constant growth rate of 4% p.a.

the right (and the flexibility) to expand capacity down the road. This stage-specific information is shown in Table 8.2. All table-related information and contents are assumed given, including the terminal values. Our real options view of this phased investment project stems from the basic notion that capacity expansion is an option and associated costs are discretionary expenditures. Viewed in this light, the investment project's value should be at least $30.85 million; clearly, TIMS Inc. is not obligated to invest in the second stage. Given this perspective, the ROV is:

$$NPV_{\text{ROV}} = NPV_{\text{DCF}}(\text{Stage 1 assets}) + \text{expansion option value}$$
$$\text{expansion option value} = \text{call}[\text{Stage 2 assets}]$$
$$NPV_{\text{ROV}} = NPV_{\text{DCF}}(\text{Stage 1 assets}) + \text{call}[\text{Stage 2 assets}]$$

Toward computing the expansion option value as a call on Stage 2 assets, we can employ the B–S & M (1973) call option formula, which is given by equations (8.4) and (8.5). This requires a parameter calibration relative to the real option problem. This benchmarking is shown in Table 8.3.
In our example, the set of input parameters benchmarked in Table 8.3 imply the following parameterization values (see Table 8.2, Stage 2):

- $S(0) = PV(FCF) = \$50.28 + \$46.66 + \$346.59 = \$443.53$
- $X = \$850$
- $\text{T} = 5\,\text{years}$
- $\sigma = 20\%$ per annum (i.e., .20).
- The volatility in our example is unknown and the value indicated is essentially a guess, at this point. Note that from a statistical standpoint, the volatility cannot be estimated based on the Table 8.2 furnishings. In most applications of real options, if available data does not provide the basis for estimating the volatility, simulation is used to provide a value for the volatility parameter. In this example, a range of values for the volatility parameter is considered to provide for the sensitivity of the results.
- $r = 3.0\%$ per annum, compounded continuously.
- $\delta = 0\%$ per annum

Through employing equation (8.4), $C(S, 0) = C_0 = S(0)e^{-\delta T}\Phi(d_1) - Xe^{-rT}\Phi(d_2)$

$$C_0 = \$443.53e^{-(0)(5)}\Phi(-.8954) - \$850e^{-(.03)(5)}\Phi(-1.34269)$$
$$\Phi(-.8954) = .18257 \quad \& \quad \Phi(-1.34269) = .08969$$
$$C_0 = \$16.557$$

The value of Stage 2 assets, viewed as an option, is $16.557 million. Therefore,

$$NPV_{\text{ROV}} = NPV_{\text{DCF}}(\text{Stage 2 assets}) + \text{call}[\text{Stage 2 assets}]$$
$$NPV_{\text{ROV}} = \$30.85 + \$16.557 = \$47.407$$

TABLE 8.2 Cash Flow Projections and Net Present Value by Investment Stage (the information shown is given)

Year	**0**	**1**	**2**	**3**	**4**	**5**	**6**	**7**	**8**
Stage 1 Investment:									
Free cash flow		32.49	41.16	(0.74)	95.13	150.02	38.59	42.13	42.64
+ Terminal value									371.25
− Investments	**(381.80)**								
Discount factor @12%	1.000	0.893	0.797	0.712	0.636	0.567	0.507	0.452	0.404
PV by year	**(381.80)**	29.01	32.81	(0.52)	60.46	85.13	19.55	19.06	167.16
NPV	**$ 30.85**								
Year	**0**	**1**	**2**	**3**	**4**	**5**	**6**	**7**	**8**
Stage 2 Investment:									
Free cash flow						-	99.24	103.14	104.40
+ Terminal value									753.75
− Investments						**(850.00)**			
Discount factor @12%						0.567	0.507	0.452	0.404
PV by year						(482.31)	50.28	46.66	346.59
NPV	**$ − 38.78**								
Year	**0**	**1**	**2**	**3**	**4**	**5**	**6**	**7**	**8**
Combined Stages:									
Free cash flow		32.49	41.16	(0.74)	95.13	150.02	137.83	145.27	147.04
+ Terminal value									1,125.00
− Investments	**(381.80)**					**(850.00)**			
Discount factor @12%	1.000	0.893	0.797	0.712	0.636	0.567	0.507	0.452	0.404
PV by year	**(381.80)**	29.01	32.81	(0.52)	60.46	(397.19)	69.83	65.71	513.76
PV	**$ − 7.93 million**								

TABLE 8.3 Real Option Parameterization

B–S & M (1973) Option Pricing Model Parameters	Notation	Real Options Parameter Mapping and Analogy
Stock price at time zero	$S(0)$	PV[free cash flows and terminal value], discounted to time zero at the risk adjusted discount rate (RADR)
Exercise price at maturity	X	Cost of asset placement
Maturity time	T	Time until the asset is placed
Volatility or standard deviation of stock price returns	σ	Volatility of phase 2 assets (standard deviation of cash flow returns).
Risk free rate	r	Risk free rate
Dividend yield	δ	1. Convenience yield[13] 2. In some settings, it can be interpreted as an opportunity cost (e.g., waiting) 3. In other settings, it can be interpreted as an adjustment for a constant inflation rate

Compared to, $NPV_{\text{DCF}} = \$ -7.93$, an $NPV_{\text{ROV}} = \$47.41$ clearly suggests that the manufacturing facility should be constructed. In addition, note too that, if the opportunity cost of waiting to expand capacity was, for instance, 1.50% per annum (i.e., $\delta = 1.50\%$) then the resulting NPV_{ROV} would decrease ($C_0 = \$11.293$). This provides a sense of how δ enters the model. In the following subsection, the implications and aptness of the use of B–S & M (1973) model in our real options context is probed and discussed in some detail.

8.3.2 APTNESS

Consider the set of building block assumptions formalizing the B–S & M (1973) model. Which ones were violated when the model was illustrated above in a real options context? Of these violations, how many are indisputable? Which ones are reasonably debatable and why? In sum, how fitting is the ROV application of the B–S & M (1973) model? What firm generalizations can be drawn from these

[13] The notion of *Convenience Yield* plays an important role in the analysis of real options. The concept is described by Dixit and Pindyck (1994); Amram and Kulatilaka (1999); and Copeland and Antikarov (2001). Brennan and Schwartz (1985) also provide an excellent perspective. Briefly, the convenience yield from a commodity or physical good represents the economic benefit that is realized from its ownership due to local and short-term shortages. This benefit accrues to the owner of the commodity not the owner of a contract for future delivery of the commodity (much the same way that the dividend paid out accrues to the owner of a stock and not the owner of an option contract on the stock). In most applications, the convenience yield is modeled as a constant proportion to the spot price, net of holding costs. This constancy assumption is largely a modeling simplification and convenience yield can also be represented as a stochastic process, typically a mean reverting process (see Schwartz 2002).

questions? Toward answering these questions, it should be noted that while stocks and options are traded securities in financial markets, projects and the real options within them, are not. The notion that the underlying source of uncertainty is a traded asset is an essential ingredient in the construction of a perfect hedge and risk neutral valuation approach. In this sense, violations of this standard can lead to serious misvaluations.

8.3.2.1 Distribution. In the B–S & M (1973) option pricing model, the underlying source of uncertainty (the stock price) is assumed to be a log-normally distributed variable. In our ROV example, the underlying source of uncertainty corresponds to the free cash flows from operations. Yet, there are no statistical or anecdotal evidence that the free cash flows, as shown in Stage 2, are log-normally distributed. In applying the B–S & M (1973) model to the real problem, the assumption of log-normality has not been fully addressed. While the accuracy of the results in this sense may be concerning, the accuracy of the forecasted cash flows (which is also utilized by the DCF method) may be more of an issue. In the current context, forecasted cash flows have been adapted to the option model where their forecast accuracy is deemed more concerning.

8.3.2.2 Discounting. The DCF methodology is structured so that all cash flows and costs are discounted at same risk adjusted rate. In the ROV approach demonstrated, the risky cash flows are (correctly) discounted at the risk adjusted rate. However, the expansion cost is discounted at the riskless rate. This contradicts the basic principle of uniform discounting in DCF. Note that in the B–S & M (1973) model, the expected present value of the exercise price is discounted at the risk free rate of return. Unless there are compelling reasons for discounting the expansion cost at the riskless rate, this may suggest that we have molded the problem to fit the solution. The only rational justification (for discounting the expansion cost at the riskless rate) would be that the risks associated with the capacity expansion are largely void of market-related or systematic risk. In other words, the capacity expansion costs reflect purely idiosyncratic risks, and for valuation purposes it may be viewed as a "zero beta asset". Permit-related setbacks due to bureaucratic or administrative blunders and construction delays resulting from weather-related problems are examples of this situation. However, if this is not the case, then using the model becomes disputable.

8.3.2.3 Volatility. The volatility parameter value used in our example was essentially a guess in order to demonstrate the B–S & M (1973) model in a real options context. Recall that in the model, the volatility parameter σ is assumed to be a given constant (i.e., homoscedastic volatility term). In practice, two distinct approaches can be adopted in estimating this parameter notwithstanding the trivial and somewhat unlikely case of an analogous "twin" project where an "implied volatility" may be obtained (see Hull 2008). The first approach is placeMonte Carlo simulation where simulated free cash flows constitute a corresponding probability distribution. This provides for an empirically based estimate of the distribution's moments, mean and variance to say the least. The second approach involves creating a portfolio of similar and independent manufacturers of hard drives. The standard deviation of return on this portfolio is a sound proxy as a volatility estimate. In general, accurate estimates of the volatility parameter are hard to come by and comparative statics (sensitivity analysis) over a range of parameter values can help in assessing model robustness or sensitivity to the parameter in question. Furthermore, the assumption of homoscedasticity may not be practically plausible, which leads to yet another violation of the model's basic assumptions. Nonetheless, the fact that volatility is explicitly build into the analysis as a value driver is an important feature of the ROV methodology. This concept is, however, ignored by DCF methods.

An indirect implication of the B–S & M (1973) model's application is that it is possible to duplicate the value of the capacity expansion option through a replicating portfolio of traded securities. Had the free cash flows resulting from the capacity expansion option been directly linked to the prices of traded assets or commodities, the replicating portfolio implication would seem more reasonable. For instance, had TIMS, Inc. been a mining company contemplating a copper extraction venture, constructing a perfect hedge through *futures* market would seem credible. However, hard drives are not traded assets. Therefore, the tacit assumption of risk neutrality seems unfitting. Unlike their financial counterparts, real options are not traded, thus limiting the range of their applicability.

In concluding this section, also note that from a modeling perspective per se, the probability that the capacity expansion option is exercised is approximately 0.09 (i.e., $\Phi(d_2) = \Phi(-1.34269) = .08969$ as shown in page 209) as opposed to the DCF model where the exercise probability is implicitly taken to be 1.0. This makes intuitive sense, on a standalone basis Stage 2 is a negative NPV proposition. Therefore, there is no stern financial incentive (in likelihood terms) to exercise the option. Given the free cash flows and expansion cost estimates, the low exercise probability actually adds to the value of capacity expansion option. In particular, given the low probability of "being in the money," the value of not having to expand capacity is conditionally higher.

The ROV approach demonstrated here provides insight to the basic intricacies of methodology described. As noted earlier, the approach also attaches an exercise probability to the capacity option and provides a value for the corresponding operational flexibility afforded, all within the confines of the assumptions that formalize the approach. In particular, the value of the investment with flexibility (to expand capacity) is given by:

$$V_0^{\text{Flex}} = V_0^{\text{Static}} + \text{flexibility premium}$$
$$\text{flexibility premium} = V_0^{\text{Flex}} - V_0^{\text{Static}} = \$47.41 - (-\$7.93) = \$55.34$$

This essentially implies that the value of this investment is largely derived from the management's flexibility to (wait and to) expand capacity down the road. To see this in a particularly simple and more familiar context, consider a decision tree view of the same problem, as shown in Figure 8.1. Suppose that the current forecast of time $t = 5$ accounts for twenty states of the nature, wherein four states justify exercising the capacity expansion option. Hence, there is a 20% likelihood

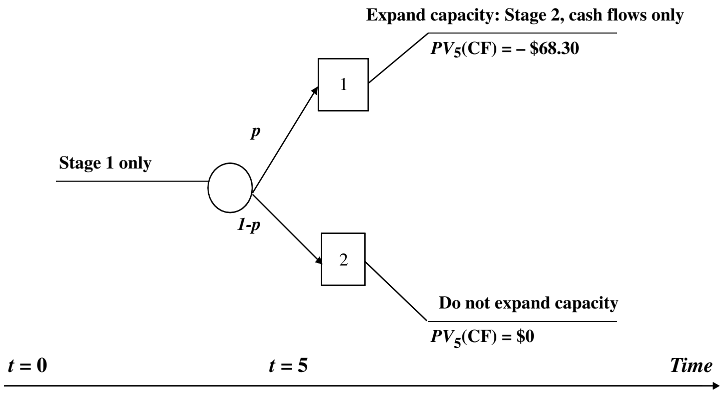


FIGURE 8.1 **Decision Tree: In the fifth year, contingent on the demand outcome (e.g., favorable with probability, $p = .20$), the company can either expand or choose not to expand capacity.**

($p = \frac{4}{20} = .20$) that TIMS, Inc expands capacity at $t = 5$, where we assume that the corresponding cash flows are defined by Table 8.2 (Stage 2 cash flows). Effectively, in Figure 8.1, we show a corresponding decision tree (DT) analogy to the expansion option value. That is,

$$NPV_{\text{DT}} = NPV_{\text{DCF}}(\text{Stage 1 assets}) + \mathbb{E}[PV(\text{expansion option value})_{\text{DT}}]$$
$$\mathbb{E}[PV(\text{expansion option value})_{\text{DT}}] = (.20)[-\$68.30(1.12)^{-5}] + (.80)[\$0]$$
$$NPV_{\text{DT}} = \$30.85 + (.20)[-\$68.30(.5674)] = \$23.10$$

Note that $-\$68.30$ is the present value of the capacity expansion at $t = 5$, that is, $\$99.24(1.12)^{-1} + \$103.14(1.12)^{-2} + (\$104.40 + \$753.75)(1.12)^{-3} -$ $850.0. Note further that as the number of scenarios where exercising capacity expansion increases (i.e., as $p \to 1.0$), $NPV_{\text{DT}} \to NPV_{\text{DCF}}$.

8.4 Conclusion

The real options (contingent claims) approach offers clear and insightful advantages that justify its adoption in suitable settings. A key requisite in this sense is market completeness: facilitating the construction of a perfect hedge and its consequent risk neutral valuation approach. That is, when a perfect hedge is feasible, the approach precludes the need for establishing a discount rate since by offsetting market-based risks, the risk free rate is the appropriate choice. The concept of risk neutrality is central not only to correctly implementing the methodology but also to obtaining economically meaningful results. The more unattainable the perfect hedge, the more economically meaningless the results are if the methodology is applied. Generally, in projects where underlying source(s) of uncertainty reflect traded assets or commodities for which future contracts trade establishing a perfect hedge is feasible (for related discussions, see Dixit and Pindyck 1994, Amram and Kulatilaka 1999, and Copeland and Antikarov 2001). Precisely for this reason, ROV applications have typically focused on mining-, farming-, harvesting-, and extraction-based problems where the uncertainty in the input or output prices (e.g. copper, crude oil, timber) can be "perfectly" hedged in financial markets.

We should add that the methodology's effectiveness is further highlighted when projects are characterized by high degree of uncertainty and significant operating flexibility. In this context, irreversibility and future strategic considerations in terms of other follow on opportunities make the approach particularly suitable. By drawing a comparison to two other commonly employed valuation techniques, DCF and DTA, we have also reflected on the advantages while identifying the relative shortcomings of the ROV approach. One clear outcome of this comparative view is that all commonly employed methods of valuation have their respective advantages and drawbacks as each is founded in its own economically based set of assumptions. When feasible, the decision maker is better served by applying all methods, contrasting the results to assess the reason for their differences.

We have clearly highlighted the advantages of an ROV approach. Perhaps the most limiting facet of the ROV approach is that for the majority of operational problems a perfect hedge is not readily afforded. Consequently, the requisite risk neutrality argument will not apply. We note that for most production- and service-based applications, operating cash flows are not directly linked to the prices of traded securities and commodities. Many finished goods, assembled products, subassemblies, and a host of manufactured items reflect goods and products that are not traded in financial markets. Yet, recent related literature has in part addressed this shortcoming by suggesting a partial equilibrium approach to the problem. Essentially, if the project's cash flows are reasonably (highly) correlated with the prices of a financially traded security, then Merton's (1973) intertemporal capital asset pricing approach can be employed to address the investment decision problem. Constantinides (1978), McDonald and Siegel (1985), Birge (2000), and Kamrad, Siddique and Ernst (2010) are the suggested references in this regard.

REFERENCES

Amram M., Kulatilaka N. 1999. Real Options: Managing Strategic Investment in an Uncertain World. *Harvard Business School Press*, Boston, MA.

Andreau S.A. 1990. A capital budgeting model for product-mix flexibility. *Journal of Manufacturing Operations Management*, **3**:5–23.

Birge J.R. 2000. Option models for incorporating risk into linear capacity planning models. *Manufacturing and Service Operations Management*, **2**(1):19–31.

Black F., Scholes M. 1973. The pricing of options and corporate liabilities. *Journal of Political Economics*, **81**(3):637–654.

Bollen NPB. 1999. Real options and product life cycles. *Management Science*, **45**(5):670–684.

Brealy R.A., Myers S.C. 2000. *Principles of Corporate Finance*. Sixth Edition, Irwin McGraw-Hill Publishing.

Brennan M.J., Schwartz E.S. 1985. Evaluating natural resource investments. *Journal of Business*, **58**:135–157.

Burnetas A., Ritchken P. 2005. Option pricing with downward sloping demand curves. *Management Science*, **51**(4):566–580.

Copeland T., Antikarov V. 2001. *Real options: A Practitioner's guide*. Texere LLC, New York.

Cox J.C., Ross S. 1976. The valuation of options for alternative stochastic processes. *Journal of Financial Economics*, **3**(1):145–166.

Cox J.C., Rubinstein M. 1985. *Options Markets.* Prentice–Hall, Englewood Cliff, NJ.

Dixit A., Pindyck R. 1994. *Investment Under Uncertainty*. Princeton, New Jersey: Princeton University Press.

Gaur V., Seshadri S. 2005. *Hedging inventory risk through market investments*, Manufacturing and Service Operations Management, **7**(2):103–120.

He H., Pindyck R. 1992. Investments in flexible production capacity. *Journal of Economic Dynamics and Control*, **16**:575–599.

Hodder J.E., Triantis A.J. 1993. Valuing flexibility: An impulse control framework. *Annals of operations research*, **45**:109–130.

Huchzermeier A., Cohen M. 1996. Valuing operational flexibility under exchange rate risk. *Operations Research*, **44**(1):100–113.

Hull J.C. (2008). *Options, Futures, and Other Derivatives*. Pearson Prentice-Hall,

Jarrow R.A., Rudd A. 1983. *Option Pricing*. Irwin Publishing, Homewood, IL.

Kamrad B., Ernst R. 2001. An economic model for mining and manufacturing ventures with output yield uncertainty. *Operations Research*, **49**(5):690–699.

Kamrad B., Lele S. 1998. Production, operating risk and market uncertainty: A valuation perspective on controlled policies. *IIE Transactions*, **30**(5):455–468.

Kamrad, B., Ord K. 2006/2007. Market risk and process uncertainty in production operations. *Naval Research Logistics*, **53**:627–640.

Kamrad B., Siddique A. 2004. Supply contracts, profit sharing, switching and reaction options. *Management Science*, **50**(1):64–82.

Kamrad B., Siddique A., Ernst R. 2010. *Partial equilibrium in risk based production decisions*. Georgetown University, McDonough School of Business, working paper.

Kogut B. 1991. Joint ventures and the options to expand and acquire. *Management Science*, **37**(1):19–33.

Kogut B., Kulatilaka N. 1994. Operating flexibility, global manufacturing and the option value of a multinational network. *Management Science*, **40**(1):123–139.

Kouvelis P., Axarloglou K., Sinha V. 2001. Exchange rates and the choice of ownership structure of production facilities. *Management Science*, **47**(8):123–139.

Kouvelis P., Tian Z. 2010. Flexible capacity investments and product mix: Optimal decision and value of postponement options. Working Paper/under review, Olin Business School. St Louis Missouri: Washington University.

Kulatila N., Perotti E.C. 1998. Strategic growth options. *Management Science*, **44**(8):1021–1031.

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

Luehrman T. 1998. *Harvard Business Review* (HBR-98404).

Merton R.C. 1973. Theory of rational option pricing. *Bell Journal of Economics and Management Science*, **4**:141–183.

Miltersen K.R., Schwartz E.S. 1998. Pricing options on commodity futures with stochastic term structures of convenience yields and interest rates. *Journal of Financial and Quantitative Analysis*, **33**:133–59.

Myers S. Finance theory and financial strategy 2001. Interfaces, January – February, 1984. Reprinted, in *Real Options and Investment under Uncertainty*. Schwartz E.S., Trigeorgis L. editors, Cambridge Massachusetts: MIT Press.

Ryan P.A., Ryan G.P. 2002. Capital budgeting practices of the fortune 1000: How have things changed. *Journal of Business and Management*, **8**(4):355.

Spinler S., Huchzermeier A. 2006. The valuation of options on capacity cost and demand uncertainty. *European Journal of Operational Research*, **171**:915–934.

CHAPTER NINE

Financial Valuation of Supply Chain Contracts

MUSTAFA Ç. PINAR, ALPER ŞEN, AND A. GÖKAY ERÖN

We consider a single buyer—single supplier multiple period quantity flexibility contract in which the buyer has options to order additional quantities of goods in case of a higher than expected demand in addition to the committed purchases at the beginning of each period of the contract. We take the buyer's point of view and find the maximum value of the contract for the buyer by analyzing the financial and real markets simultaneously. We assume both markets evolve as discrete scenario trees. Under the assumption that the demand of the item is perfectly positively correlated with the price of a risky security traded in the financial market, we present a model to find the buyer's maximum acceptable price of the contract. Applying duality theory of linear programming, we obtain a martingale expression for the value of the contract. Finally, an experimental study is presented to illustrate the impacts of various parameters on the values of the contract and the option.

Handbook of Integrated Risk Management in Global Supply Chains, First Edition. Edited by Panos Kouvelis, Lingxiu Dong, Onur Boyabatli, and Rong Li.

9.1 Introduction

In an effort to reduce mismatches between supply and demand, many companies have recently been looking into their relationships with their supply chain partners. These relationships are formally structured through supply contracts which specify the terms of a binding agreement between a seller and a buyer, whether the buyer is a manufacturer purchasing raw materials from a supplier, a wholesaler purchasing finished goods from a manufacturer or a retailer purchasing goods from a wholesaler. The terms in a supply contract include pricing and discounts, minimum quantities and flexibility terms, delivery terms, lead times, and quality and product return policies. Proper specification of these terms to align the incentives and coordinate the supply chain has become an important subject in operations management practice and literature.

Many supply contracts focus on the use of contractual flexibility in terms of purchased quantities in better handling the demand uncertainty at the end customer level. Various flexibility terms can be employed to transfer the risk to a party that can absorb it at a lower cost. For example, under a *backup agreement* for fashion goods, the buyer commits to a total quantity over a single selling season. The seller delivers a prespecified portion of the total quantity at the start of the season. The buyer may buy additional units up to the remaining commitment during the season and pays a penalty for any portion of the committed units that he did not purchase. Such contracts are used by major apparel and catalog companies and their suppliers (Eppen and Iyer 1997). In a similar contract with *return* or *buyback* terms, the buyer may return unsold items at a credit to the seller (Pasternack 1985). Return provisions are commonplace in distribution of many goods including books, newspapers, recorded music, and computer hardware and software (Padmanabhan and Png 1995). In *quantity–flexibility* contracts, the buyer first provides a forecast to the seller before season and is able to order within prespecified limits of this forecast during the season. These contracts are used heavily in high tech industry (Bassok et al. 1997, Tsay 1999). In *capacity reservation* contracts, the buyer reserves a portion of the seller's capacity at a reservation price, which he can later use by paying an execution price (Barnes-Schuster et al. 2002, Serel et al. 2001). Capacity reservation contracts are important as (1) other major contracts mentioned above can be shown to be their special cases (see Barnes-Schuster et al. 2002 and Cachon and Lariviere 2001) and (2) they mimic the (call) options in financial markets, which received enormous attention since the seminal article by Black and Scholes (1973). Note that in addition to contract terms discussed above, there are other provisions in supply contracts to entice the buyer to take more risk and order more. These include price protection in personal computers (Lee et al. 2000), markdown money in consumer goods (Tsay 2002), and revenue sharing terms in video rental industries (Cachon 2005). The now broad operations management literature on supply contracts generally studies the benefits of different contractual terms on buyer and/or seller, the factors that moderate these benefits and the ability of these terms to coordinate the supply chain under various settings and assumptions. Two excellent reviews of this literature are provided in Cachon (2003) and Tsay et al. (1998).

In this chapter, we develop a formal approach for valuation of flexible supply contracts under demand uncertainty. In particular, we consider the valuation of the capacity reservation or *option* contracts in which the buyer has the option (but not obligation) to purchase additional quantities up to the capacity reservation during the season, in addition to possible firm commitments (*forwards*) to be delivered before or during the season. Several researchers addressed the issue of valuation or pricing of these contracts in operations management literature. In Cheng et al. (2006), the authors model the price negotiation process as a Stackelberg game, where the seller is the leader and determines the option/exercise prices and the buyer is the follower and chooses an initial firm commitment and the number of option contracts to purchase in a single period setting. In Barnes-Schuster et al. (2002), the authors use a two-period model with correlated demand between periods and determine the seller's optimal pricing and production decisions. Here, the seller determines the wholesales prices for firm commitments in two periods as well as the optimal option/exercise prices. In Li et al. (2009), the authors assume that the buyer is privately informed about the market price and demand, and show that this asymmetry in information may lead to a different seller valuation of these contracts. In Ritcken and Tapiero (1986), the authors determine the conditions under which the options provide benefits over inventory building when the market demand and spot price are stochastic and correlated for a risk-averse decision maker with quadratic utility. In all four models, however, the analysis is carried out in the absence of financial markets. Ignoring the impact of financial markets is common not only in supply contracts literature, but also in other areas of operations management research. This is despite the extensive use of financial instruments by real sector companies when managing their operations. For example, according to a recent study, 60.3% of 7,319 nonfinancial firms worldwide use financial derivatives (Bartram et al. 2009).

The main contribution in this chapter is the formal integration of financial markets in valuation of flexible supply contracts under uncertain demand. This is important for three reasons: First, the financial markets may provide another means (in addition to the flexibility terms provided in the supply contract) to hedge against the inventory risk faced by individual players in the supply chain. Second, the true valuation (net of the value obtained by trading in financial markets) of flexibility terms can be obtained. Third, public (and perhaps advance) information regarding the financial markets can be used for better specification of contract terms and increasing the profitability under these terms due to possible correlation between financial and real market evolutions.

Several papers study the impact of financial markets on operational decisions. Gaur and Seshadri (2005) study the problem of hedging the inventory risk in a single period newsvendor model when the demand is correlated with the price of a financial asset. The authors show how to construct optimal hedging transactions and among other things, show that hedging increases the order quantity of a risk averse newsvendor. Burnetas and Ritchken (2005) study the role of option contracts when the retail price is a linear function of a stochastic market size factor and total quantity released to the market (purchased goods plus goods available through exercised options). It is assumed that there is a traded security that spans the uncertainty in the demand curve and a riskless bond that leads to a complete market and valuations that are independent of risk preferences of the buyer and seller. The conditions under which the options benefit the seller and buyer are investigated. Ding et al. (2007) study the integrated operational and financial hedging decisions of a global firm selling in two markets with demand and exchange rate uncertainties. It is shown that the operational and financial hedging strategies are tightly interlinked and the lack of use of financial hedges can have significant effect on supply chain structural decisions such as the location and the number of production facilities to satisfy global demand. Caldentey and Haugh (2009) study the performance of a supply chain that consists of seller and buyer in a newsvendor setting. The buyer purchases a single product from the seller and resells it at the retail market at a stochastic retail price. It is assumed that the returns in a financial market and the retail price are dependent. It is also assumed that the buyer is budget constrained. Under a flexible contract, the seller offers a menu of wholesale contracts based on the particular evolution of the financial market until a certain time before the season and the buyer commits to an order quantity. Under a flexible contract with hedging, the buyer is able to continuously trade in the financial market before he places his order. It is shown that while the seller always prefers the flexible contract with hedging, the choice depends on the model parameters for the buyer. Finally, Chen and Parlar (2007) study the value of a put option for a risk averse vendor. The payoff of the put option in this model is contingent on the realized value of the demand.

In this paper, we use a stochastic programming approach for the financial valuation of flexible supply contracts. In this regard, we follow the approach in King (2002), who develops a stochastic programming formulation for pricing contingent claims (options) in the discrete time, discrete state case. King (2002) shows that the absence of arbitrage in the hedging problem is equivalent to the existence of a probability measure that makes the price process a martingale in the dual problem. He also shows that in complete markets the dual problem determines the unique valuation operator (equivalent risk neutral martingale measure). Delft and Vial (2004) use a similar stochastic programming approach in evaluation of supply contracts. Their model, however, does not include any financial instruments and thus does not capture the important impact of financial markets.

As in King (2002), we assume that financial and real markets evolve as discrete scenario trees. We further assume that there is a perfect correlation between the demand and the price of a risky asset traded in the financial market, which implies that the scenario trees of the markets coincide. This assumption is partially validated by Gaur and Seshadri (2005) who show that there is significant correlation between the year-to-year same-store sales growth in 60 large U.S. retailers and the same period returns on the S&P 500 index. In our model, the buyer borrows by short selling stocks in the financial market to acquire the contract and an investment portfolio. He closes the short position in later periods by transactions in financial markets and proceeds of operations in the real market. Thus, our approach allows us to find the maximum price that the buyer should accept to pay for the contract by studying the financial and real markets simultaneously.

The organization of the chapter is as follows: In Section 9.2, we review the stochastic process governing the security prices. Furthermore, we introduce financial markets and the basic concepts (arbitrage and martingales) of our analysis. In Section 9.3, the real market is introduced and the relation between the financial and real markets is described. The assumptions and notations are listed, and the model is described in detail. In Section 9.4, we apply duality theory of linear programming to obtain a martingale expression for the value of the contract. In Section 9.5, we conduct a numerical study to analyze the parameters affecting the valuation of the buyer, and give some managerial insights. We conclude in Section 9.6 with a short summary.

9.2 Review of Financial Markets, Arbitrage, and Martingales

A financial market is a mechanism that allows people to buy and sell financial securities. Throughout the chapter we assume as in King (2002) that all random quantities are supported on a finite probability space $(\Omega, \mathcal{F}, P)$ whose atoms ω are sequences of real valued vectors (security prices and payments) over the discrete time periods $t = 0, 1, \ldots, T$. In addition, we assume that the market evolves as a discrete scenario tree. In the scenario tree, the partition of probability atoms $\omega \in \Omega$, which are generated by matching path histories up to time t corresponds one-to-one with nodes $n \in N_t$ at level t in the tree. The root node $n = 0$ corresponds to trivial partition $N_0 = \Omega$, and the leaf nodes $n \in N_T$ correspond one-to-one with the probability atoms $\omega \in \Omega$.

As shown in Figure 9.1, in the scenario tree, every node $n \in N_t$ for $t = 1, \ldots, T$ has a unique parent node denoted by $a(n) \in N_{t-1}$, and every node $n \in N_t$, $t = 0, 1, \ldots, T - 1$ has a nonempty set of child nodes denoted by $C(n) \subset N_{t+1}$. The tree evolution described Figure 9.1 is more general than a

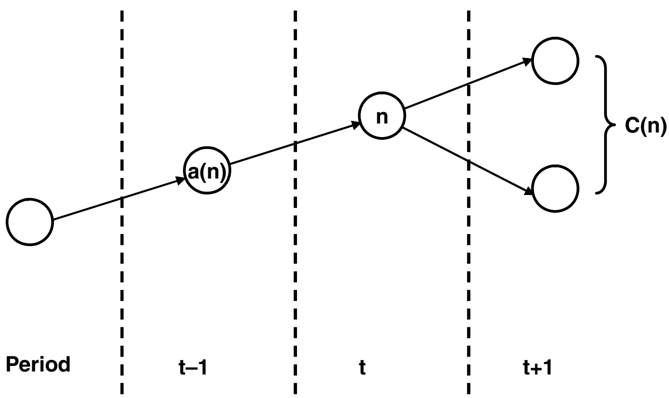


FIGURE 9.1 Financial Market Scenario Tree

recombinant binomial tree evolution usually used in introductory options pricing analysis. In a binomial tree with recombination (i.e., when a node can have more than one parent) and two assets (one riskless and one risky asset) the market is complete, and the portfolio strategies are naturally chosen to be path-independent. However, in incomplete markets it is known that path-independent strategies can be suboptimal. Therefore, to implement path-dependent strategies, the nonrecombinant tree evolution model (where each node has a unique parent) described in this chapter is more suitable. The examples given in the chapter are binomial but are nonrecombinant, hence suitable for incomplete markets. Therefore, the models used in the chapter are general as discussed in Edirisinghe et al. (1993).

The probability distribution P assigns positive weights p_n to each leaf node $n \in N_T$ in such a way that $\sum_{n \in N_T} p_n = 1$. Each intermediate level node in the tree receives a probability mass equal to the combined mass of the paths passing through it:

$$p_n = \sum_{u \in C(n)} p_u \qquad \forall n \in N_t, \qquad t = T-1, \ldots, 0$$

The ratios p_u/p_n, $u \in C(n)$, are the conditional probabilities that the child node u occurs given that the parent node $n = a(u)$ has occurred.

The function $X : \Omega \rightarrow \mathbb{R}$ is a real-valued random variable if $\{\omega : X(\omega) \leq r\} \in \mathcal{F}\, \forall r \in \mathbb{R}$. Let X be a real–valued random variable. X can be lifted to N_t if it can be assigned a value on each node of N_t that is consistent with its definition on Ω (King 2002). This kind of random variable is said to be measurable with respect to the information contained in the nodes of N_t. A stochastic process $\{X_t\}$ is a time indexed collection of random variables such that each X_t is measurable with respect to N_t. The expected value of X_t is uniquely defined by

$$E^P[X_t] := \sum_{n \in N_t} p_n X_n$$

The conditional expectation of X_{t+1} on N_t is a random variable taking values over the nodes $n \in N_t$:

$$E^P[X_{t+1} | N_t] := \sum_{u \in C(n)} \frac{p_u}{p_n} X_u$$

The market consists of $J+1$ traded securities indexed by $j = 0, 1, \ldots, J$ with prices at node n given by the vector $S_n = \left(S_n^0, \ldots, S_n^J\right)$. We suppose one of the securities, say security 0, always has strictly positive values at each node of the scenario tree. This security that corresponds to the risk free asset (bond) in the classical financial valuation framework is chosen to be *numéraire*. Introducing the discount factors $\beta_n = 1/S_n^0$, we define the discounted security prices relative to the numéraire and denote it by $Z_n = \left(Z_n^0, \ldots, Z_n^J\right)$ where $Z_n^j = \beta_n S_n^j$ for $j = 0, 1, \ldots, J$. Note that, $Z_n^0 = 1$ in any state n.

The amount of security j held by the investor in state $n \in N_t$ is denoted by θ_n^j. The value of the portfolio discounted with respect to the numéraire in state n is

$$Z_n \cdot \theta_n := \sum_{j=0}^{J} Z_n^j \theta_n^j$$

An *arbitrage* is a sequence of portfolio holdings that begins with a zero initial value, makes self-financing portfolio transactions and attains a non-negative value in each future state, while in at least one terminal state it attains a strictly positive value with positive probability.

The condition of self-financing portfolio transactions in the following equation states that the funds available for investment at state n are restricted to the funds generated by the price changes in the portfolio held at state $a(n)$.

$$Z_n \cdot \theta_n = Z_n \cdot \theta_{a(n)} \qquad n > 0$$

The following optimization problem is used to find an arbitrage:

$$\begin{aligned} \max \quad & \sum_{n \in N_T} p_n Z_n \cdot \theta_n \\ & \text{s.t.} \\ Z_0 \cdot \theta_0 &= 0 \\ Z_n \cdot \left[\theta_n - \theta_{a(n)}\right] &= 0, \qquad \forall n \in N_t,\ t \geq 1 \\ Z_n \cdot \theta_n &\geq 0, \qquad \forall n \in N_T \end{aligned}$$

The solution that yields a positive optimal value can be turned into an arbitrage as shown by Harrison and Pliska (1981). On the other hand if no arbitrage is possible, the price process is called an arbitrage-free market price process.

Martingale properties needed for our study are formalized in the following definition.

Definition 9.1 *If there exists a probability measure $Q = \{q_n\}_{n \in N_t}$ such that $Z_t = E^Q\left[Z_{t+1} \mid N_t\right]$, for all $t \leq T - 1$, then the vector process $\{Z_t\}$ is called a vector-valued martingale under Q, and Q is called a martingale probability measure* (*MPM*) *for the process.*

We further need the following definition.

Definition 9.2 *A discrete probability measure $Q = \{q_n\}_{n \in N_t}$ is said to be equivalent to a discrete probability measure $P = \{p_n\}_{n \in N_t}$ if $q_n > 0$ exactly when $p_n > 0$.*

The key link between arbitrage and martingales is the following theorem (c.f. Theorem 9.1 of King (2002)).

THEOREM 9.1 *The discrete state stochastic vector process $\{Z_t\}$ is an arbitrage-free market price process if and only if there is at least one probability measure Q equivalent to P under which $\{Z_t\}$ is a martingale.*

9.3 A Model for Financial Valuation of Supply Chain Contracts

We consider a general single buyer–single supplier contract where the buyer is an intermediary between the market and the supplier. He buys the finished products from the supplier and sells them to customers at the end market at a fixed market price that is exogenously specified. The buyer and the supplier sign a multiple period quantity flexibility contract, in which the buyer has options to place further orders in case of a higher than expected demand in addition to the committed purchases at the beginning of each period of the contract.

We assume that the demand of the customers for the finished products evolves as a discrete scenario tree. The nodes of the scenario tree represent the state of the discrete state stochastic process at a given period. The arcs correspond to the probabilistic transitions from one node at a given period to another node at the next period. As represented in Figure 9.2, there exists exactly one arc leading to a node, while there may be many arcs emanating from a node. As in the financial market scenario tree we denote the nodes obtained by the arcs emanating from node n, $n \in N_t$ for $t = 0, \dots, T-1$ by $C(n) \subset N_{t+1}$, and the unique node that gives rise to node n, $n \in N_t$ for $t = 1, \dots, T$ by $a(n) \in N_{t-1}$.

Now, consider a periodic review inventory problem with horizon T. The decisions made by the buyer at the beginning of the horizon are as follows. The buyer orders Q_t units to be delivered in period t for $t = 1, \dots, T$ at a unit purchase price of p_t. We refer to Q_t as firm orders. In addition, the buyer purchases options from the supplier, which give him an opportunity to purchase additional

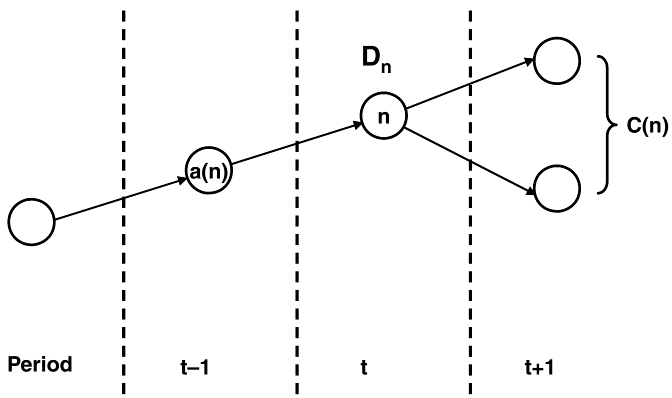


FIGURE 9.2 Demand Market Scenario Tree

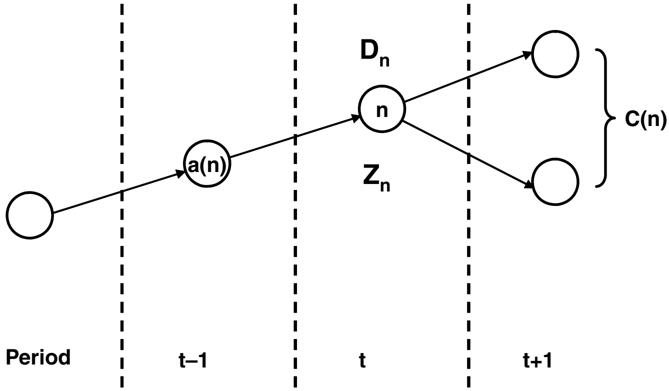


FIGURE 9.3 Financial and Demand Market Scenario Tree

units later by paying an exercise price. We assume that one option gives the buyer a right to purchase one additional unit of product, and this additional unit is delivered at the beginning of the next period. We further assume that the number of options exercised by the buyer at each node n, $n \in N_t$ for $t = 1, \ldots, T-1$ is denoted by m_n and is bounded above by a constant M. In each state n, $n \in N_t$ for $t = 1, \ldots, T-1$, after observing the actual demand of node n, the buyer decides whether to exercise options or not. Exercise price in period t is e_t.

In each period t for $t = 1, \ldots, T-1$, excess demand is assumed to be backlogged to the next period at a unit shortage cost s_t. However, at the end of the horizon, shortage is not allowed. In addition, in each period t, $t = 1, \ldots, T$, excess inventory is carried to the next period at a unit holding cost of h_t.

We assume that demand forecast for the item is perfectly correlated with the price of a risky security traded in the financial market implying that the scenario tree of the financial market and the demand market coincide as shown in Figure 9.3.

Before moving on to the mathematical formulation of the model, we summarize the notation and assumptions relevant to the model.

9.3.1 NOTATION

Decision Variables

V^M: Contract value with M options

Q_t : Firm order to be delivered in period t

θ_n : The vector amount of securities held at node n

m_n : Number of options exercised at node n

I_n^+ : Positive inventory at the end of node n

I_n^- : Negative inventory at the end of node n

I_n : Net inventory at the end of node n

Parameters

- M: Maximum number of options that can be exercised at node n
- r_t: Sales price of finished product at the end market in period t
- p_t: Purchase price of unit firm order Q_t in period t
- h_t: Unit holding cost for finished products in period t
- s_t: Unit stock-out cost for finished products in period t
- Z_n: The vector of security prices at node n
- D_n: Demand at node n
- e_t: Unit price for an option exercised in period t

9.3.2 ASSUMPTIONS

1. The demand forecast for the item is perfectly correlated with the price of an underlying security traded in the financial markets.
2. In the financial market, the price process $\{Z_t\}$ is an arbitrage-free market price process. This is equivalent to the existence of a martingale probability measure Q for the price process $\{Z_t\}$.
3. At each state n, $n \in N_t$ for $t = 1, \ldots, T-1$, the buyer is allowed to exercise at most M options and the options exercised are delivered at the beginning of period $t+1$.
4. In the real market, in period t for $t = 1, \ldots, T-1$ excess demand is backlogged and excess inventory is carried to the next period. However, at the end of the horizon, shortage is not allowed.
5. The backorders are met at the present price.
6. To avoid trivial cases, it is assumed that the sales price r_t is greater than the purchase price p_t and the stock-out cost s_t is greater than the holding cost h_t in period t for $t = 1, \ldots, T$.
7. Initial inventory is assumed to be zero, and we have no salvage value at the end of the horizon.
8. The buyer is assumed to be risk-neutral.
9. The firm orders (Q_t) and the number of options to be exercised (m_n) are assumed to take real values.

9.3.3 THE MODEL

The aim of the model is to find the maximum value (V^M) that the buyer is willing to pay for the contract. Hence, the objective is to maximize the value of the contract. Since the portfolio of the buyer is empty before borrowing money, and the money borrowed (by shorting stocks or bonds) at the beginning of the horizon is used to acquire the contract and acquire the stock or bond to later trade in the financial market, the portfolio of stocks, bonds and the value of the contract must add up to *zero*, which constitutes $Z_0 \cdot \theta_0 + V^M = 0$ (constraint [9.1] below).

The portfolio value at each node n, $Z_n \cdot \theta_n$, is composed of the portfolio value of parent node $a(n)$, $Z_n \cdot \theta_{a(n)}$, and the cash flow generated in the real market at node n denoted by F_n. Therefore, the following equation describes the self-financing nature of portfolio transactions:

$$Z_n \cdot \theta_n = Z_n \cdot \theta_{a(n)} + F_n$$

or,

$$Z_n \cdot \left(\theta_n - \theta_{a(n)}\right) = F_n$$

Denote $\theta_n - \theta_{a(n)}$ by $\Delta\theta_n$ then we have

$$Z_n \cdot \Delta\theta_n = F_n$$

With above specifications, our model, referred to as $(P1)$, can be formulated as follows.

$$\max V^M$$

s.t.

$$Z_0 \cdot \theta_0 + V^M = 0 \tag{9.1}$$

$$Z_n \cdot \Delta\theta_n + r_1 I_n^- + p_1 Q_1 + e_1 m_n + h_1 I_n^+ + s_1 I_n^- = r_1 D_n \qquad \forall n \in N_1 \tag{9.2}$$

$$Z_n \cdot \Delta\theta_n + r_t \left(I_n^- - I_{a(n)}^-\right) + p_t Q_t + e_t m_n + h_t I_n^+ + s_t I_n^- = r_t D_n \qquad \forall n \in N_t, t = 2, \ldots, T-1 \tag{9.3}$$

$$Z_n \cdot \Delta\theta_n - r_T I_{a(n)}^- + p_T Q_T + h_T I_n = r_T D_n \qquad \forall n \in N_T \tag{9.4}$$

$$Z_n \cdot \theta_n \geq 0 \qquad \forall n \in N_T \tag{9.5}$$

$$Q_1 - I_n = D_n \qquad \forall n \in N_1 \tag{9.6}$$

$$I_{a(n)} + Q_t + m_{a(n)} - I_n = D_n \qquad \forall n \in N_t, t = 2, \ldots, T \tag{9.7}$$

$$I_n - I_n^+ + I_n^- = 0 \qquad \forall n \in N_t, t = 1, \ldots, T-1 \tag{9.8}$$

$$I_n \geq 0 \qquad \forall n \in N_T \tag{9.9}$$

$$m_n \leq M \qquad \forall n \in N_t, t = 1, \ldots, T-1 \tag{9.10}$$

$$m_n \geq 0 \qquad \forall n \in N_t, t = 1, \ldots, T-1 \tag{9.11}$$

$$Q_t \geq 0 \qquad t = 1, \ldots, T \tag{9.12}$$

$$I_n^+ \geq 0 \qquad \forall n \in N_t, t = 1, \ldots, T \tag{9.13}$$

$$I_n^- \geq 0 \qquad \forall n \in N_t, t = 1, \ldots, T-1 \tag{9.14}$$

Constraint (9.2) implies that F_n for $n \in N_1$ is the revenue in period 1, which is composed of the amount of the product sold at a unit sales price of r_1, minus the expenditure in period 1, which is the firm order at a unit purchase price of p_1, the amount of options exercised to be used in the second period at a unit exercise price of e_1, the positive inventory at a unit cost of h_1 and the backorder amount at a unit cost of s_1:

$$F_n = r_1\left(D_n - I_n^-\right) - \left(p_1 Q_1 + e_1 m_n + h_1 I_n^+ + s_1 I_n^-\right) \qquad \forall n \in N_1$$

Constraint (9.3) states that F_n for $n \in N_t$, $t = 2, \ldots, T-1$ is the revenue in period t, $t = 2, \ldots, T-1$, that is, the demand at node n plus the backorder amount at node $a(n)$ minus the shortage at node n at a unit sales price of r_t, minus the expenditure in period t, $t = 2, \ldots, T-1$, that is, the firm order, the number of options exercised in period t to be used in period $t+1$, the positive inventory and the backorder amount at unit prices of p_t, e_t, h_t and s_t:

$$\begin{aligned} F_n = r_t &\left(D_n - I_n^- + I_{a(n)}^-\right) \\ &- \left(p_t Q_t + e_t m_n + h_t I_n^+ + s_t I_n^-\right) \forall n \in N_t,\ t = 2, \ldots, T-1 \end{aligned}$$

Constraint (9.4) ensures that F_n for $n \in N_T$ is the revenue in the last period, which is the demand at node n plus the backorder amount coming from parent node $a(n)$ at a unit sales price of r_T since shortage is not allowed in the last period, minus the expenditure, which is the firm order at a unit purchase price p_T plus the positive inventory held at node n at a unit cost of h_T since in the last period options cannot be exercised and shortage is not allowed:

$$F_n = r_T\left(D_n + I_{a(n)}^-\right) - \left(p_T Q_T + h_T I_n\right) \qquad \forall n \in N_T$$

Constraint (9.5) guarantees that the value of the portfolio in the terminal states are non-negative. This is needed to assure that the buyer has repaid fully the initial debt from short positions.

Constraints (9.6), (9.7), (9.8), and (9.9) are the inventory balance constraints. Constraint (9.6) implies that in the first period the net inventory at each state n, $n \in N_1$ is equal to the firm order for period 1 minus the demand at that node since there is no backorder to cover or positive inventory carried from the previous period.

Constraint (9.7) states that in period t, $t = 2, \ldots, T$ the net inventory at each state n, $n \in N_t$ is equal to the sum of the net inventory of the parent node $a(n)$, the firm order of period t and the number of options exercised in period $t-1$ to be delivered in period t minus the demand at state n. The reason is that except the first period, the buyer is allowed to carry positive or negative inventory from the previous periods. Furthermore, the buyer has an opportunity to use options bringing as many additional units as the number of options exercised.

Constraint (9.8) implies that in period t, $t = 1, \ldots, T-1$, the net inventory at any node is equal to positive inventory minus the negative inventory at that node. However, the net inventory in the last period is simply the positive inventory, since shortage is not allowed at the end of the horizon. This is guaranteed in constraint (9.9).

Constraint (9.10) shows the flexibility of the buyer. It states that at any node that the buyer is allowed to exercise options which is all the periods except the last period, he is permitted to exercise at most M options.

The resulting optimization model is a linear programming problem, which can be efficiently solved by off-the-shelf optimization software.

9.4 Dual Formulation

This section analyzes the problem discussed in Section 9.3 through an equivalent dual formulation. We first examine the financial constraints in the dual corresponding to the decision variables θ_n for $n \in N_t$, $t = 0, \ldots, T$. The first step in calculating the dual is to assign dual variables to each constraint in the model. We assign q_n as dual variables for all the nodes of the financial constraints (9.1–9.4), and w_n for the non-negativity constraint of the portfolio in the terminal nodes, that is, constraint (9.5), $\forall n \in N_T$.

Firstly, the dual constraint corresponding to the decision variable V^M, that is the value of the contract, is:

$$q_0 = 1 \tag{9.15}$$

Next, the dual constraint corresponding to θ_n, $n \in N_t$ for $t = 0, \ldots, T-1$ is the martingale condition:

$$q_n Z_n = \sum_{u \in C(n)} q_u Z_u \qquad n \in N_t, t = 0, \ldots, T-1 \tag{9.16}$$

The dual constraint corresponding to the decision variables θ_n for $n \in N_T$ is:

$$\left(q_n + w_n\right) Z_n = 0 \qquad n \in N_T$$

Since the first component $Z_n^0 = 1$ for all states n we have:

$$q_n + w_n = 0 \qquad n \in N_T$$

In addition, by the non-negativity of the portfolio in the terminal positions:

$$w_n \leq 0 \qquad n \in N_T$$

Finally, combining the above two constraints, one has the following constraint in the dual:

$$q_n \geq 0 \qquad n \in N_T \tag{9.17}$$

We assign dual multipliers y_n to the inventory balance constraints (9.6) and (9.7), $\forall n \in N_t$, $t = 1, \ldots, T$, k_n to constraint (9.8), $\forall n \in N_t$, $t = 1, \ldots, T-1$, and f_n to the flexibility constraint (9.10), $\forall n \in N_t$, $t = 1, \ldots, T-1$. The dual constraint corresponding to the firm orders Q_t is:

$$\sum_{n \in N_t} p_t q_n + y_n \geq 0 \quad t = 1, \ldots, T \tag{9.18}$$

The constraint in the dual arising from the number of options exercised (i.e. m_n), $n \in N_t, t = 1, \ldots, T-1$ is:

$$e_t q_n + f_n + \sum_{u \in C(n)} y_u \geq 0 \quad n \in N_t, t = 1, \ldots, T-1 \tag{9.19}$$

The dual constraint corresponding to the net inventory at state n, $n \in N_t, t = 1, \ldots, T-1$ is:

$$-y_n + \sum_{u \in C(n)} y_u + k_n = 0 \quad n \in N_t, t = 1, \ldots, T-1$$

Reformulating the above constraint, one obtains:

$$k_n = y_n - \sum_{u \in C(n)} y_u \quad n \in N_t, t = 1, \ldots, T-1$$

The constraint in the dual arising from the positive inventory at state n, $n \in N_t, t = 1, \ldots, T-1$ is:

$$h_t q_n - k_n \geq 0 \quad n \in N_t, t = 1, \ldots, T-1$$

and The dual constraint associated with the negative inventory at state n, $n \in N_t, t = 1, \ldots, T-1$ is:

$$(r_t + s_t)\, q_n - r_{t+1} \sum_{u \in C(n)} q_u + k_n \geq 0 \quad n \in N_t, t = 1, \ldots, T-1$$

Replacing k_n by $y_n - \sum_{u \in C(n)} y_u$ one has the following constraints in the dual corresponding to, respectively, positive and negative inventory at state n, $n \in N_t, t = 1, \ldots, T-1$:

$$h_t q_n - y_n + \sum_{u \in C(n)} y_u \geq 0 \quad n \in N_t, t = 1, \ldots, T-1 \tag{9.20}$$

$$(r_t + s_t)\, q_n - r_{t+1} \sum_{u \in C(n)} q_u + y_n - \sum_{u \in C(n)} y_u \geq 0 \quad n \in N_t, t = 1, \ldots, T-1 \tag{9.21}$$

Finally, the dual constraint corresponding to the net inventory at the terminal positions (which is also the positive inventory since shortages are not allowed in the last period) is:

$$h_T q_n - y_n \geq 0 \quad n \in N_T \tag{9.22}$$

Therefore, the dual program, which we refer to as $(D1)$, is as follows:

$$\begin{aligned} \min \quad & \sum_{t=1}^{T} \sum_{n \in N_t} D_n \left(r_t q_n + y_n \right) + M \sum_{t=1}^{T-1} \sum_{n \in N_t} f_n \\ & \text{s.t.} \\ (15-22) & \\ & f_n \geq 0 \qquad n \in N_t, t = 1, \ldots, T-1 \end{aligned}$$

The basic theorem of linear programming states that problem $(P1)$ has an optimal solution if and only if the dual $(D1)$ does too, and both optimal values are equal. Furthermore, it follows again from the theory of linear programming that problem $(P1)$ has an optimal solution if and only if it is feasible and bounded. Moreover, $(P1)$ is bounded if and only if there exists at least one probability measure Q under which the price process $\{Z_t\}$ is martingale, and there exists y_n and f_n satisfying (9.18–9.22).

Now, assume the financial market is arbitrage-free, and let $\mathcal{M}$ denote the set of probability measures Q making the stock price process a martingale. Then, we can summarize our findings above in the result below.

Theorem 9.2 *The maximum value that the buyer will accept to pay for the contract is:*

$$\min_{Q \in \mathcal{M}} \left\{ \sum_{t=1}^{T} \sum_{n \in N_t} D_n \left(r_t q_n + y_n^* \right) + M \sum_{t=1}^{T-1} \sum_{n \in N_t} f_n^* \right\}$$

where y^* and f^* are the optimal solution of the following linear program that we refer to as $(D2)$:

$$\min \quad \sum_{t=1}^{T} \sum_{n \in N_t} D_n y_n + M \sum_{t=1}^{T-1} \sum_{n \in N_t} f_n$$

s.t.

$$\sum_{n \in N_t} y_n \geq - \sum_{n \in N_t} p_t q_n \qquad t = 1, \ldots, T \tag{9.23}$$

$$f_n + \sum_{u \in C(n)} y_u \geq -e_t q_n \qquad n \in N_t, t = 1, \ldots, T-1 \tag{9.24}$$

$$y_n - \sum_{u \in C(n)} y_u \leq h_t q_n \qquad n \in N_t, t = 1, \ldots, T-1 \tag{9.25}$$

$$y_n - \sum_{u \in C(n)} y_u \geq r_{t+1}$$

$$\sum_{u \in C(n)} q_u - (r_t + s_t)\, q_n \qquad n \in N_t, t = 1, \ldots, T-1 \tag{9.26}$$

$$y_n \leq h_T q_n \qquad n \in N_T \tag{9.27}$$

$$f_n \geq 0 \qquad n \in N_t, t = 1, \ldots, T-1. \tag{9.28}$$

We first note that when the financial market is complete and arbitrage-free, the set $\mathcal{M}$ is a singleton in which case, it suffices to solve $(D2)$ in Theorem 9.2 to solve the dual problem, given the unique martingale measure, Q^* say.

From Theorem 9.2, we can also make the following observation.

OBSERVATION 9.1 *If $f_n^* = 0$, an increase in the value of M does not have any effect on the value of the contract since*

$$M \sum_{t=1}^{T-1} \sum_{n \in N_t} f_n^* = 0$$

This actually means that the buyer is flexible enough to exercise as many options as he wants even before an increase in the value of M, that is, the primal constraints corresponding to f_n for $n \in N_t$, $t = 1, \ldots, T-1$ are all nonbinding.

9.5 Experimental Study

In this section, we explore the parameters that moderate the benefits that the buyer receives from options in supply contracts. The value of $M = \mu$ options (often referred to as *option value* in the rest of the chapter) available in a supply contract can be determined by subtracting the value of the contract when $M = 0$ from the value of the contract when $M = \mu$. This is necessary, as the model we study has an operating profit even when the use of option is not allowed. The value of a contract is found by solving the linear program $(P1)$ in Section 9.3.3. For simplicity, we first conduct all the analysis in a two-period model and consider the binomial tree shown in Figure 9.4. A three-period model is considered when

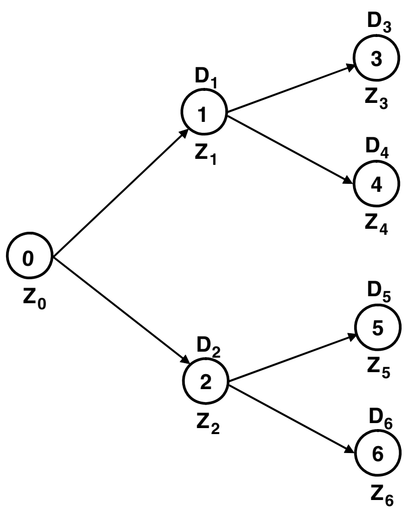


FIGURE 9.4 Two–Period Binomial Tree with $p = 1/2$

need arises. We assume that there is only one risky security and one riskless asset in the financial market.

From Figure 9.4, we have

$$N_0 = \{0\}, \qquad N_1 = \{1, 2\}, \qquad N_2 = \{3, 4, 5, 6\}$$

$$a(1) = 0, \qquad a(2) = 0$$

$$a(3) = 1, \qquad a(4) = 1$$

$$a(5) = 2, \qquad a(6) = 2$$

$$Z_n = \left(Z_n^0, Z_n^1\right) \qquad n = 0, \ldots, 6$$

Note that Z_n^0 denotes the price of the riskless asset, and Z_n^1 denotes the price of the risky security.

In order to observe the effect of volatility of stock prices, the stock prices are chosen in such a way that the average price remains constant in all periods:

$$Z_0^1 = \left(Z_1^1 + Z_2^1\right)/2 = \left(Z_3^1 + Z_4^1 + Z_5^1 + Z_6^1\right)/4$$

Demands at each node are also set so that the average values remains constant in all periods to observe the impact of demand volatility:

$$(D_1 + D_2)/2 = (D_3 + D_4 + D_5 + D_6)/4$$

Under the above specifications, the values of the parameters and the corresponding decision variables in a base case are represented in Table 9.1. The value of the contract (V^M) is 146.7857 and the value of the option is $V^M - V^0 = 482.30 - 416.68 = 65.62$. Notice that the buyer takes a short position in the stock, the proceeds of which are used to finance the contract and the purchase of bonds.

Throughout the analysis, graphs are plotted by taking the sample size of the parameters large enough to recognize a general pattern. Solid lines represent the value of the contract, and the dashed lines represent the value of the option.

TABLE 9.1 Parameters and Decision Variables in Base Case

Parameters	Decision Variables	n	Z_n^0	Z_n^1	D_n	θ_n^0	θ_n^1	F_n	m_n
	$Q_1 = 45$	0	10	15		45.902	−62.754		
$r_t = 20$	$Q_2 = 20$	1	12	20	45	−13.064	−26.875	10.0	35
$p_t = 12$	$V^0 = 416.68$	2	12	10	25	19.767	−38.393	−70.0	0
$h_t = 1.5$	$V^M = 482.30$	3	14.4	25	55			860.0	
$s_t = 2.5$	$I_2^+ = 20$	4	14.4	5	30			322.5	
$M = 100$	$I_4 = 25$	5	14.4	22	40			560.0	
$e = 10$	$I_6 = 25$	6	14.4	8	15			22.5	

9.5.1 CASE 1: EFFECT OF NUMBER OF OPTIONS

The buyer is allowed to purchase options from the supplier at the beginning of the horizon to later exercise and obtain additional units. The buyer, however, is not fully flexible to adjust order quantities to the observed demands. At each state $n, n \in N_t, t = 1, \ldots, T - 1$, he is allowed to exercise at most M options. Thus, the value of M plays an important role in determining the value of the contract and the option. As shown in Figure 9.5, the values of the contract and the option are unchanged as long as the buyer is flexible enough to exercise the amount used in the base case. However, decreasing the value of M to an amount lower than the amount of options exercised in the base case decreases the values of the contract and the option.

9.5.2 CASE 2: EFFECT OF EXERCISE PRICE

The buyer can use options to obtain additional units only after paying an exercise price. Thus, the price that the buyer pays to exercise options affects the values of the option and the contract. Obviously, an increase (decrease) in exercise price leads to a decrease (increase) in the value of the contract and option value. This is shown in Figure 9.6 for our problem.

9.5.3 CASE 3: EFFECT OF PURCHASE PRICE

At the beginning of the horizon, the buyer orders Q_t units at a unit purchase price of p_t to be delivered in period t, $t = 1, \ldots, T$. Hence, the value of the purchase price has an effect on the value of the contract and the option. Table 9.2 shows the impact of purchase price in periods 1 and 2 on stock and option values. As expected, the value of the contract is nonincreasing in purchase prices, while the option value is nondecreasing. When the purchase price in period 1 (p_1) is too low at 6, the buyer places only firm orders in period 1, and does not exercise any options. This leads to zero option value. When p_1 is increased to 8, the buyer reduces the firm orders, and exercise options in the high demand

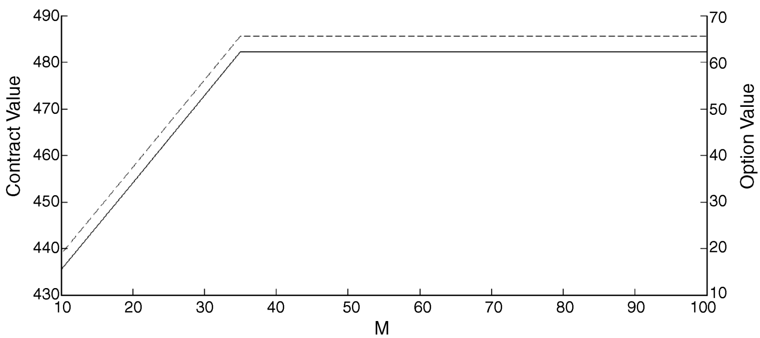


FIGURE 9.5 **Contract and Option Values vs Number of Options (M)**

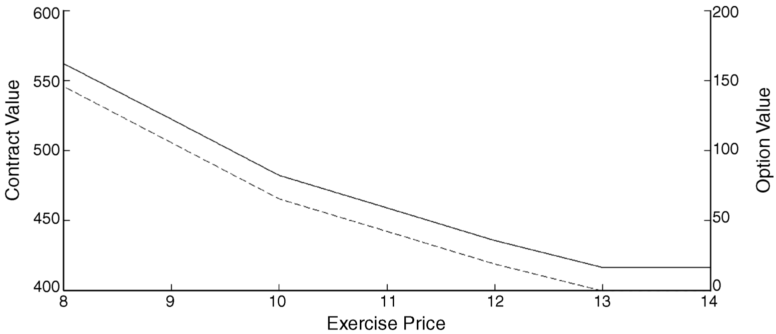


FIGURE 9.6 **Contract and Option Values vs Exercise Price**

scenario leading to a positive option value. At $p_1 = 10$, the buyer places a firm order of only 45 units in period 1 and uses options in period 2, regardless of the demand realization. Further increases in p_1 has no impact on option value since it does not impact the way the buyer uses options for period 2. When the purchase price in period 2 (p_2) is too low at 9, the buyer places only firm orders for periods 1 and 2 leading to again zero option value. As p_2 increases to 10, the buyer uses a mixture of firm orders and options in the high demand scenario, increasing the value of the options. When p_2 increases to 12, the buyer uses only options for period 2, which leads to a further increase in the value of the options.

9.5.4 CASE 4: EFFECT OF DEMAND VOLATILITY

More volatile demand leads to more mismatches between the supply and demand for the buyer. Since the options are used to correct mismatches of period 1 and

TABLE 9.2 **Impact of Purchase Price with $M = 100$**

p_1	p_2	Q_1	Q_2	m_1	m_2	V^0	V^M	$V^M - V^0$
6	12	100	0	0	0	756.26	756.26	0.00
8	12	65	0	35	0	640.63	589.59	51.04
10	12	45	0	55	20	491.68	557.30	65.62
12	12	45	20	35	0	416.68	482.30	65.62
14	12	45	20	35	0	341.68	407.30	65.62
16	12	0	65	35	0	297.38	363.00	65.62
12	9	25	75	0	0	533.48	533.48	0.00
12	10	45	20	35	0	493.07	510.08	17.01
12	11	45	20	35	0	454.87	496.19	41.32
12	12	45	20	35	0	416.68	482.30	65.62
12	13	45	0	55	20	378.48	482.30	103.82
12	14	45	0	55	20	340.29	482.30	142.01

TABLE 9.3 Impact of Demand Volatility with $M = 100$

D_1	D_2	D_3	D_4	D_5	D_6	Q_1	Q_2	m_1	m_2	V^0	V^M	$V^M - V^0$
40	30	55	30	40	15	40	0	55	30	421.26	468.13	46.87
45	25	55	30	40	15	45	20	35	0	416.68	482.30	65.62
50	20	55	30	40	15	50	10	45	0	412.07	496.47	84.40
45	25	50	35	40	15	45	0	50	20	409.80	466.05	56.25
45	25	55	30	40	15	45	20	35	0	416.68	482.30	65.62
45	25	60	25	40	15	45	20	40	0	423.55	498.55	75.00

to minimize the possible mismatch of period 2 by adjusting orders in accordance with observed demands, they are more valuable when the demand is more volatile. Table 9.3 shows the impact of demand volatility on contract and option values. The first half is regarding the volatility in period 1 and the second half is regarding the volatility in period 2. Both parts show that the options become more valuable as the volatility increases.

9.5.5 CASE 5: VOLATILITY OF STOCK PRICES

In order to analyze the impact of stock prices on contract and option values, we vary the volatility of the stock prices while keeping the mean of the stock prices constant throughout the horizon, that is:

$$Z_0^1 = \left(Z_1^1 + Z_2^1\right)/2 = \left(Z_3^1 + Z_4^1 + Z_5^1 + Z_6^1\right)/4$$

The value of the option corresponding to different values of stock prices are summarized in Table 9.4. We assume that other parameters take their base case values.

The first row of Table 9.4 (base case) shows that assuming the stock price pattern above, the buyer makes 62.754 short sales of stocks at the beginning of the horizon. The portfolio of stocks in node 1 and node 2, respectively, are −26.875 and −38.393. This implies that the buyer has paid back part of the debt and has 26.875 and 38.393 remaining stocks to pay in node 1 and node 2, respectively. The value of the option is 65.62.

TABLE 9.4 Impact of Stock Volatility with $M = 100$

Z_0^1	Z_1^1	Z_2^1	Z_3^1	Z_4^1	Z_5^1	Z_6^1	θ_0	θ_1	θ_2	V^0	V^M	$V^M - V^0$
15	20	10	25	5	22	8	−62.754	−26.875	−38.393	416.68	482.30	65.62
15	19	11	25	5	22	8	−70.285	−26.875	−38.393	464.91	505.93	41.02
15	20	10	28	2	24	6	−55.970	−20.673	−29.861	389.22	454.84	65.62
15	21	9	28	2	24	6	−50.853	−20.673	−29.861	356.19	438.22	82.03
15	19	11	28	2	24	6	−63.646	−20.673	−29.861	436.84	477.86	41.02

First, we keep the stock prices in period 2 constant and analyze the effect of stock prices in period 1. We first observe that as the volatility of the stock prices in period 1 decreases, the value of the option also decreases, since the demand is perfectly correlated with the price of a risky security and it is period 1 in which the options are exercised.

Next, we investigate the case where the stock prices in period 1 are unchanged. The stock prices in period 2 do not have any impact on the value of the option. This is due to the fact that in period 2 (the terminal position), the buyer cannot exercise any options. However, the stock prices in period 2 impact the portfolio of stock in period 1, as the buyer needs to cover all his short sales and forms his portfolio in period 1 by considering the stock prices in the next period. This result is stated in the following observation.

Observation 9.2 *The stock prices in period* 2 *do not impact the value of the option, whereas they impact the portfolio of stock in period* 1.

9.5.6 CASE 6: EFFECT OF INTEREST RATE ON THE RISKLESS ASSET

The value of the option corresponding to different interest rates are summarized in Table 9.5. The values of the other parameters are taken as in the base case. The first row of Table 9.5 (base case) shows that if the interest rate on the riskless asset is 20%, the value of the option is 65.62. The buyer places 45 and 20 units firm orders for period 1 and 2 respectively. In node 1 the buyer exercises 35 options. As the interest rate on the riskless asset decreases, the buyer can make short sales of bonds in larger quantities and exercise more options with the cash borrowed to meet the demand in case of higher than expected demand. This leads to use of less firm orders and more options, and thus option value increases.

Thus far, all the cases were analyzed in a two-period setting. Since the analysis of the remaining parameters requires a higher dimensional model, we now extend our model to three periods, and consider the binomial tree shown in Figure 9.7.

From Figure 9.7, we have,

$$N_0 = \{0\}, \qquad N_1 = \{1, 2\}, \qquad N_2 = \{3, 4, 5, 6\}$$

$$N_3 = \{7, 8, 9, 10, 11, 12, 13, 14\}$$

$$a(1) = a(2) = 0$$

TABLE 9.5 Impact of Interest Rate with $M = 100$

Interest rate(%)	Q_1	Q_2	m_1	m_2	V^0	V^M	$V^M - V^0$
20	45	20	35		416.68	482.30	65.62
10	45		55	20	286.66	458.66	172.00
0	25		75	40	111.20	452.45	341.25
25	45	20	35		469.22	497.22	28.00

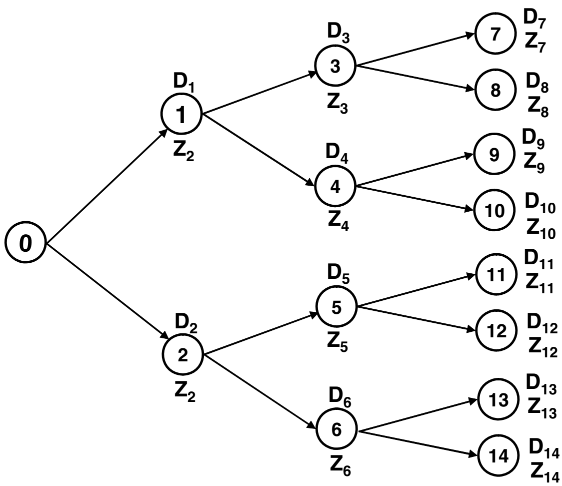


FIGURE 9.7 **Three-Period Binomial Tree**

$$a(3) = a(4) = 1, \quad a(5) = a(6) = 2$$
$$a(7) = a(8) = 3, \quad a(9) = a(10) = 4$$
$$a(11) = a(12) = 5, \quad a(13) = a(14) = 6$$
$$Z_n = \left(Z_n^0, Z_n^1\right) \quad n = 0, \ldots, 14$$

As in the two-period case, we chose a base case such that the average stock prices and demand values remain constant over the periods:

$$Z_0^1 = \sum_{i=1}^{2} Z_i^1/2 = \sum_{i=3}^{6} Z_i^1/4 = \sum_{i=7}^{14} Z_i^1/8,$$
$$\sum_{i=1}^{2} D_i/2 = \sum_{i=3}^{6} D_i/4 = \sum_{i=7}^{14} D_i/8$$

The values of the parameters and the corresponding decision variables in the base case are shown in Table 9.6.

We now study the effect of following parameters on the value of the contact and option.

9.5.7 CASE 7: EFFECT OF SALES PRICE

The buyer sells the finished products at the end market to the customers at unit sales prices of r_1, r_2, and r_3 in period 1, 2, and 3, respectively. Figure 9.8 shows the impact of first period sales price (r_1) on the contract and option values. The

TABLE 9.6 Parameters and the Decision Variables in Base Case

Parameters	Decision Variables	n	Z_n^0	Z_n^1	D_n	θ_n^0	θ_n^1	F_n
		0	10.00	15		10.440	−60.223	
$r_t = 20$	$Q_1 = 40$	1	11.00	20	45	−71.663	−29.702	−292.5
$p_t = 12$	$V^0 = 353.70$	2	11.00	10	25	−23.363	−23.300	−2.5
$h_t = 1.5$	$V^M = 799.10$	3	12.10	25	55	23.479	−53.750	550.0
$s_t = 2.5$	$m_1 = 60$	4	12.10	5	30	−4.132	−80.625	562.5
$M = 100$	$m_3 = 65$	5	12.10	22	40	31.528	−76.786	−512.5
$e_t = 10$	$m_4 = 10$	6	12.10	8	15	28.174	−107.500	−50.0
	$m_5 = 75$	7	13.31	30	65			1300.0
	$m_6 = 35$	8	13.31	20	40			762.5
		9	13.31	8	35			700.0
		10	13.31	4	20			377.5
		11	13.31	25	50			1500.0
		12	13.31	18	25			962.5
		13	13.31	10	35			700.0
		14	13.31	5	10			162.5

contract value obviously increases with r_1, as the end product sales bring more revenue to the buyer. The option value, on the other hand, is decreasing in r_1, as the buyer will ensure that the demands are satisfied with firm orders in period 1 with a higher probability, thus will exercise fewer options at the beginning of period 1. After a threshold value of r_1, the option value does not change, since the buyer does not use any options to satisfy period 1 demand after this threshold.

9.5.8 CASE 8: EFFECT OF HOLDING COST

Inventory holding cost rate has a direct effect on contract and option values. Figure 9.9 shows the impact of holding cost rate in period 1 (h_1). As h_1 increases,

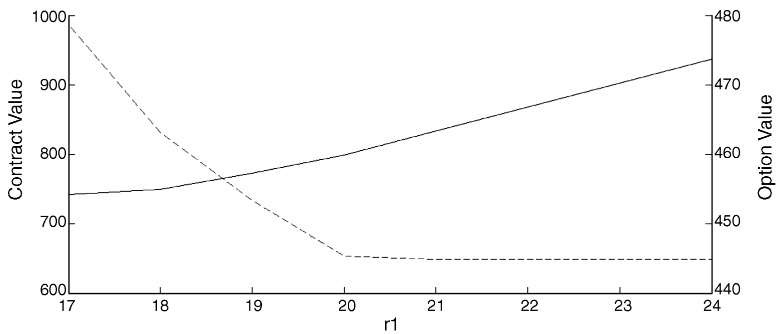


FIGURE 9.8 Contract and Option Values vs Sales Price of Period 1

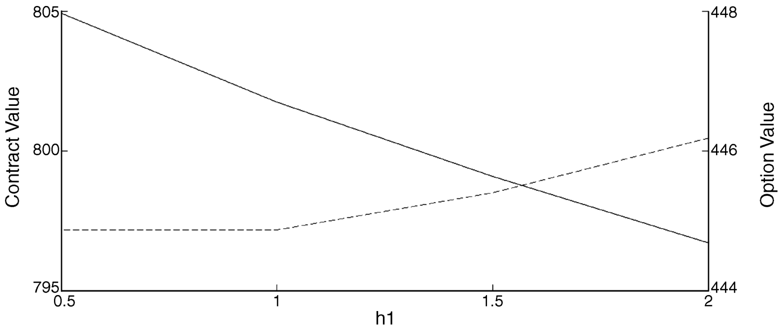


FIGURE 9.9 **Contract and Option Values vs Holding Cost of Period** 1

the contract value obviously decreases, since the buyer incurs more cost for excess inventory. On the other hand, the option value increases since the buyer prefers to use options instead of placing firm orders and carrying inventory to satisfy demand.

A similar result is shown in Figure 9.10, where the inventory holding cost rates in periods 2 and 3 are changed.

9.5.9 CASE 9: EFFECT OF STOCK-OUT COST

If the buyer cannot satisfy the demand in a given period, he incurs a stock-out cost. Therefore, the stock-out costs has an immediate effect on the value of the contract and the option value. Figure 9.11 shows the impact of the period 1 stock-out cost on the contract and option value. As expected, the contract value decreases with an increase in stock-out cost. However, we also see a decrease in the option value. This happens since the buyer places more firm orders in order to avoid stock-outs and thus needs less options when the stock-out

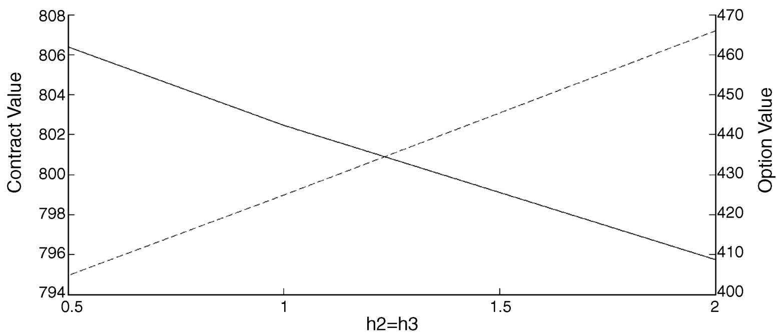


FIGURE 9.10 **Contract and Option Values vs Holding Cost of Period 2–3**

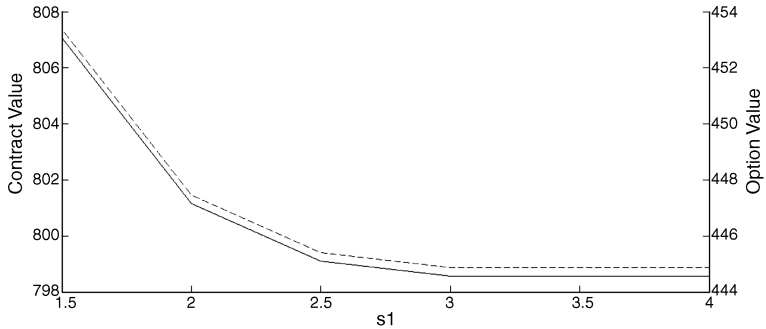


FIGURE 9.11 **Contract and Option Values vs Stock-out Cost of Period** 1

cost is high (Remember that the options that are exercised lead to deliveries next period).

9.6 Conclusion

In this chapter, we considered a general single buyer—single supplier quantity flexibility contract with options for multiple periods. We investigated the problem of the buyer of the contract under the assumption that the markets evolve as discrete scenario trees, and gave a linear optimization model to compute the maximum acceptable price of the contract for the buyer by analyzing both financial and real markets. Since the relationship between various parameters in the model can be quite intricate to analyze, an experimental study for parameter shifts, which are too complicated for analysis, was also presented. The model is flexible enough to accommodate incomplete financial markets, markets with frictions (transaction costs and taxes), and risk aversion attitudes of the buyer. Future research will extend the model to the aforementioned cases as well as the interesting case of partial correlation of the demand and the price of the risky security.

REFERENCES

Barnes-Schuster D., Bassok Y., Anupindi R. 2002. Coordination and flexibility in supply contracts with options. *Manufacturing and Service Operations Management*, **4**:171–207.

Bartram S.M., Brown G.W., Fehle F.R. 2009. International evidence on financial derivatives usage. *Financial Management*, **38**:185–206.

Bassok Y., Srinivasan R., Bixby A., Wiesel H. 1997. Design of component supply contracts with commitment revision flexibility. *IBM Journal of Research and Development*, **41**:693–703.

Black F., Scholes M. 1973. The pricing of options and corporate liabilities. *Journal of Political Economy*, **81**:637–654.

Burnetas A., Ritchken P. 2005. Option pricing with downward-sloping demand curves: The case of supply chain options. *Management Science*, **51**:566–580.

Cachon G. 2003. Supply chain coordination with contracts. In: Graves C.S., de Kok T., editors. *Handbooks in Operations Research & Management Science: Supply Chain Management*. North Holland.

Cachon G. 2005. Supply chain coordination with revenue-sharing contracts: Strengths and limitations. *Management Science*, **51**:30–44.

Cachon G., Lariviere M.A. 2001. Contracting to assure supply: How to share demand forecasts in a supply chain. *Management Science*, **47**:629–646.

Caldentey R., Haugh B.M. 2009. Supply contracts with financial hedging. *Operations Research*, **57**:47–65.

Chen F., Parlar M. 2007. Value of a put option to the risk-averse newsvendor. *IIE Transactions* **39**:481–500.

Cheng F., Ettl M., Lin G.Y., Schwarz M., Yao D. 2006. Designing flexible supply chain contracts via options. Technical report. Yorktown Heights, New York: IBM Thomas J. Watson Research Center.

Delft van Ch., Vial J.Ph. 2004. A practical implementation of stochastic programming: An application to the evaluation of option contracts in supply chains. *Automatica*, **40**:743–756.

Ding Q., Dong L., Kouvelis P. 2007. On the integration of production and financial markets in global markets. *Operations Research*, **55**:470–489.

Edirisinghe C., Naik V., Uppal R. 1993. Optimal replication of options with transaction costs and trading restrictions. *Journal of Financial and Quantitative Analysis*, **28**:117–138.

Eppen G., Iyer E. 1997. Backup agreements in fashion buying—the value of upstream flexibility. *Management Science*, **43**:1469–1484.

Gaur V., Seshadri S. 2005. Hedging inventory risk through market instruments. *Manufacturing and Service Operations Management*, **7**:103–120.

Harrison J.M., Pliska S.R. 1981. Martingales and stochastic integrals in the theory of continuous time trading. *Stochastic Processes and Their Applications*, **11**: 215–260.

King, A.J. 2002. Duality and martingales: A stochastic programming perspective on contingent claims. *Mathematical Programming* **91**:543–562.

Lee H.L., Padmanabhan V., Taylor T.A., Whang S. 2000. Price protection in the personal computer industry. *Management Science*, **46**:467–482.

Li H., Ritchken P., Wang Y. 2009. Option and forward contracting with asymmetric information: Valuation issues in supply chains. *European Journal of Operational Research*, **197**:134–148.

Padmanabhan V., Png I.P.L. 1995. Returns policies: Make money by making good. *Sloan Management Review*, **37**:65–72.

Pasternack B.A. 1985. Optimal pricing and return policies for perishable commodities. *Marketing Science*, **4**:166–176.

Ritcken P.H., Tapiero C.S. 1986. Contingent claims contracting for purchasing decision in inventory management. *Operations Research*, **34**:864–870.

Serel D.A., Dada M., Moskowitz H. 2001. Sourcing decisions with capacity reservation contracts. *European Journal of Operational Research*, **131**:635–648.

Tsay A., Nahmias S., Agrawal N. 1998. Modeling supply chain contracts: A review. In: Magazine M., Tayur S., Ganeshan R., editors. *Quantitative Models for Supply Chain Management*. Kluwer Academic Publishers.

Tsay A.A. 1999. The quantity flexibility contract and supplier–customer incentives. *Management Science*, **45**:1339–1358.

Tsay A.A. 2002. Managing retail channel overstock: Markdown money and return policies. *Journal of Retailing*, **77**:457–492.

Part Three

Supply Chain Finance

CHAPTER TEN

Supply Chain Finance

PANOS KOUVELIS AND WENHUI ZHAO

We will explicitly address the financing of inventories within the supply chain, and investigate the use of bank financing versus trade credit financing. We will explore the implications of financing decisions on relevant retail ordering and supplier (wholesale) pricing decisions. We will clearly illustrate that in the presence of bankruptcy risks and related costs for the supply chain, firms' financing and operational decisions are interdependent, and the choice of financing affects the attained supply chain efficiency.

We consider a linear supply chain of a retailer (he) and a supplier (she). The newsvendor-like retailer has a single opportunity to order a product from the supplier to satisfy future uncertain demand. Both the retailer and supplier might be capital constrained and in need of short-term financing. If necessary, both are able to borrow competitively priced bank loans. We first study the wholesale price only contract when bankruptcy costs exist. We show that the retailer's optimal order quantity is a function of the wholesale price and his wealth (working capital and collateral), and it might be smaller than the traditional newsvendor order quantity due to the existence of bankruptcy costs. In the presence of bankruptcy risks, the equilibrium order quantity is smaller than the traditional equilibrium solution without capital constraints, and both the equilibrium order quantity and wholesale price increase in the retailer's wealth. We then study the supplier's trade credit contracts (w, r_s) where w is the wholesale price if the retailer pays

Handbook of Integrated Risk Management in Global Supply Chains, First Edition. Edited by Panos Kouvelis, Lingxiu Dong, Onur Boyabatli, and Rong Li.

the order up front, and r_s is the interest rate if he delays the payment. We conclude that a risk neutral supplier should always finance the retailer at rates less than or equal to the risk-free rate. The retailer, if offered an optimally structured trade credit contract, will always prefer supplier financing to bank financing. Furthermore, under supplier financing, both the supplier's profit and supply chain efficiency improve, and the retailer might also improve his profits relative to bank financing.

10.1 Introduction

Effective supply chain management integrates information, material and financial flows in a way that matches supply with demand for the firm's products and services with the needed efficiency and responsiveness according to the firm's strategy. While our research literature has placed substantial emphasis on understanding integration of information flows (such as demand forecasting, visibility of inventories and upcoming orders, information on supply capacity, and lead times) with relevant operational decisions within the supply chain (such as quantity and timing of orders, allocation of orders among suppliers, and coordinated shipments), a lot less emphasis has been placed on the understanding of implications of financial decisions on supply chain operations and achieved efficiency. Our work in this chapter addresses such needs.

We will explicitly address the financing of inventories within the supply chain. We will analyze and contrast the implications of two different forms of inventory financing: bank financing and supply financing via trade credits. Our analysis takes place within a stylized modeling context of a linear supply chain of two firms, a newsvendor-like retailer and a supplier with variable supply (ordering, production) costs. While our stylized model continues to have many elements of market perfection as the classic Modigliani-Miller (M&M) setting (refer to Modigliani and Miller 1958), we will introduce certain imperfections that break down the independence of financing and operational decisions advocated in their work. We do keep in our work market perfection such as full information shared among supply chain players, no transaction costs in borrowing money, and banks pricing their loans in a competitive manner. However, supply chain firms borrowing money do face risks in their ability to repay the loans due to demand uncertainties, and failure to do so may lead to costly bankruptcies. In the case of use of supplier-provided trade credit contracts to the retailer, we model existing best such practices (e.g., early payment discounts or open account financing) that noncompetitively price demand risks (offered percentage discounts independent of ordering quantity), as opposed to competitively by bank loan interest rates that fully reflect the order quantity and the relevant probabilistic information on the demand.

These market imperfections and transaction costs together with relevant agency behavior of the decentralized supply chain, will result in high interdependency of financing and relevant operational decisions (retailer's order quantities and supplier's wholesale prices). Our analysis clearly depicts such decision interdependencies and obtains the optimal levels for the relevant decision variables.

Limited working capital is a frequent constraint in corporate procurement decisions. In the presence of liquidity constraints, firms are in need of short-term financing to execute their procurement actions. Banks are obvious sources for short-term financing. However, in order for the bank to break even through the loan transaction, he will charge the interest rate reflecting the various risks associated with the loan, that is, the bank's interest rate will be a function of the borrower's operational decisions (e.g., the order quantity).

Competitive pricing of a loan is a standard assumption in the corporate finance literature (see Barro 1976, Bester 1994, and Tirole 2006 for more references), which is justified when the financial market is efficient and expectations are taken with respect to the risk-neutral probability measure. The use of competitively priced loan implies that the creditors' *expected* repayment from the loan (after bankruptcy costs) equals the time value of the loan. In other words, the *expected* interest rate of the bank is the risk-free interest rate.

In an effort to reduce the bank's interest rate charges, capitally constrained retailers often pledge beyond the sales receipts from the procured inventories other corporate assets (outstanding accounts receivables, existing inventories, warehouses, etc.) as collaterals for the loan (see Brealey et al. 1995 and Tirole 2006). Failure of the retailer to repay the loan might lead to a costly bankruptcy, with costs reflecting both fixed administrative costs, variable costs proportional to realized sales, and a depressed collateral value. These bankruptcy costs are discussed in detail in corporate finance literature (see Ang et al. 1982 and Tirole 2006). In the event of a bankruptcy, the bank has the right to seize all sales receipts and liquidate pledged collaterals, but only after covering the fixed administrative costs for engaging in such bankruptcy proceedings.

If the value of the pledged collateral by the retailer is higher than or equal to the loan principal and interest, the loan is fully secured by the collateral. In this case, the bank faces no bankruptcy risks of the retailer in issuing the loan, and the bank's interest rate is the risk-free rate. However, if the collateral cannot fully cover the loan principal and interest, and the realized retail sales are low (i.e., the sales receipts plus collateral value are less than the owed amount), the retailer is not able to repay the bank's loan and declares bankruptcy. To account for the relevant bankruptcy risks and costs, the bank charges an interest rate premium above the risk-free rate to make the expected interest rate the risk-free interest rate.

Within this context, we consider a supply chain where a retailer has a single opportunity to order "now" a product from a supplier to satisfy future uncertain demand. The strategic interaction between the supplier and retailer can be modeled as a Stackelberg game of wholesale price and order quantity between the supplier ("leader" in setting the wholesale price) and the retailer ("follower" in responding with the order quantity). We study both the financially constrained retailer's ordering problem in the face of uncertain demand and a fixed price wholesale contract, and the supplier's problem in setting the optimal wholesale price.

Note that in the above discussion, the retailer gets the full finance through bank loans. On the other hand, he might also consider to get short-term financing from other supply chain partners. For example, a very popular and widely witnessed practice is that the upstream supplier provides finance to the downstream retailer via trade credits. The most typical trade credit contract is that of "supplier early payment discount." It allows buyers to pay the supplier for products purchased within a given time window (e.g., 30, 60 or 90 days) without incurring financial charges. However, early payment is encouraged through a discount on the offered wholesale price. If there is no early payment discount, the trade credit practice is referred to as "open account financing." Another form of supplier financing is through financial intermediaries, such as banks, where the bank provides the loan to the buyers, the buyers pay the loan principal and the supplier pays the interest to the bank (see Zhou and Groenevelt 2007).

The prevalence of trade credit practices brings to the forefront several research questions: Why would suppliers provide financing to buyers instead of letting them use bank loans? What are the optimal trade credit contract parameters (*wholesale price* if paying early, *interest rate*, or equivalently *early payment discount rate*, if delaying payment)? What are the implications of such contracts for supply chain efficiency and the profitability of each party in the chain? We will study and answer these research questions through the previously mentioned Stackelberg game setting between the supplier and retailer.

We would like to provide a preview of our results on supplier based financing and the related analysis. We start by discussing supply chain interactions when the supplier uses a simple wholesale price contract, all needed financing of inventories is provided via competitively priced bank loans, and with general assumptions about bankruptcy costs (both fixed and variable such costs). Our analysis of the "bankrupt-prone newsvendor" shows that his optimal order quantity is not only a function of the wholesale price, but also depends on the retailer's working capital and collateral. Our analysis of the supplier's problems, or what is referred to as "selling to the bankrupt-prone newsvendor", establishes unique equilibrium prices and order quantities in the so-called retailer's bankruptcy region (e.g., positive probability for the retailer's bankruptcy) under mild assumptions. For this region, increases in the retailer's wealth lead to increased equilibrium wholesale prices, order quantities and supplier's expected profit. Furthermore, it is clear that the presence of costly bankruptcies decreases the efficiency of the supply chain under a wholesale price contract.

Next, we have been able to offer a clear supply chain theory based rationale of the attractiveness of trade credit financing of inventories. Supplier offered optimally structured trade credit contracts that underprice the retailer's demand risks and offer interest rates even below the risk-free interest rate. For example, open account financing practices with zero trade credit rate may be optimal. The low trade credit rates incentivize the retailer to order higher order quantities, for the same or higher wholesale price, and end up with higher likelihood of operating within the retailer's bankruptcy region. As a result of it, the supplier is able to obtain higher expected profits, even though it offers "cheap financing" of inventories to the retailer. The "cheap financing" has the retailer always prefer it to bank financing, and in most cases, the retailer also ends up with higher expected profits as well. Low wealth retailers might not do better under trade credit contracts and would have preferred bank financing and a wholesale price contract. However, with a supplier acting as a Stackelberg leader and dictating the contract choice, this contract alternative might not be available. Finally, trade credit contracts, even though still noncoordinating contracts, increase the supply chain efficiency as they result in larger order quantities closer to the centralized system optimal.

The remaining of our chapter has the following structure. In Section 10.2, we introduce our basic model setting, common notation, and a set of key assumptions for all of our analysis. Section 10.3 analyzes the supply chain interactions under a wholesale price contract and in the presence of bankruptcy costs. Our main results on the "bankrupt-prone newsvendor" appear in Section 10.3.2, while the "selling to the bankrupt-prone newsvendor" analysis is presented in Section 10.3.3. Our analysis and discussion on the use of trade credit versus bank financing appear in Section 10.4. The basic model setup is in Section 10.4.1, the solution of the newsvendor-like retailer's problem under trade credit contracts appears in Section 10.4.2, and the corresponding supplier's problem for optimal setting of the contract parameters is in Section 10.4.3. The implication for the supplier's profit, retailer's profit, and overall supply chain efficiency are presented in Section 10.4.4. Finally, a summary of our overall conclusions from this supply chain finance research appear in Section 10.5.

10.2 The Model Setting, Common Notation and Assumptions

We consider a supply chain where a retailer has a single opportunity to order "now" a product from a supplier to satisfy future uncertain demand. The supplier offers a take-it-or-leave-it trade credit contract/wholesale price contract to the retailer. There are two time points considered: "now" and "the end of sales season." The decisions to be made include the contract parameters, the retailer's ordering decision, his financial decisions (loan is borrowed from the bank or supplier, and the amount to borrow), the supplier's financial decisions, and the bank's interest rates when the supplier and/or the retailer approaches for a loan transaction. All decisions are made now by the retailer, supplier, and bank, based on the corresponding cash flows at the end of the sales season (expectations if cash flows are random variables).

Let y be the retailer's working capital prior to ordering and financing decisions that are made. His collateral assets pledged towards securing the short-term loan valued at the end of the sales season are denoted by x (these assets can be used to pay debt obligations in the case of bankruptcy). The exogenous parameters are r_f, the risk-free interest rate for the period from now to the end of sales season; p, the unit retail price; and c, the unit production cost. Since the salvage value of unsold items and goodwill loss for unmet demand do not change the nature of the problem, without loss of generality, they are ignored from our models. The

TABLE 10.1 Common Notation

p:	Retailer's unit retail price of the product.
c:	Supplier's unit production cost, $0 < c(1 + r_f) \leq p$.
r_f:	Risk-free interest rate for the period from now to the end of the sales season.
B:	Administrative cost of bankruptcy (fixed cost).
α_1:	Variable bankruptcy cost stated as a portion (α_1) of realized sales (i.e., $\alpha_1 p\xi$). $0 \leq \alpha_1 \leq 1$.
α_2:	Depressed collateral value in the event of a bankruptcy stated as a portion (α_2) of the originally estimated collateral value (i.e., $\alpha_2 x$), $0 \leq \alpha_2 \leq 1$.
x:	Retailer's collateral assets (valued at the end of the sales season) pledged to secure the loan.
y:	Retailer's working capital prior to purchasing and financing decisions are made.
Ω:	Retailer's wealth at the end of the sales season in terms of the working capital and collateral, $\Omega = x + y(1 + r_f)$, and we assume $\Omega > 0$.
q:	Retailer's order quantity (or equivalently, the retailer's inventory level).
w:	Supplier's unit wholesale price if the retailer pays up front.
r_b:	Bank's interest rate when the retailer borrows a bank loan, $r_b \geq r_f$.
k:	Retailer's bankruptcy threshold (i.e., the minimal demand for the retailer to repay the loan obligations (loan principal and interest)).
b:	Administrative bankruptcy threshold (i.e., the minimal realized demand to only cover the administrative (fixed) bankruptcy cost).
π:	The retailer's expected cash flows at the end of sales season.
Π:	The supplier's expected cash flows at the end of sales season.
Γ:	The bank's expected cash flows at the end of sales season.

common notation used in all the two research topics (i.e., the study of wholesale price contracts and trade credit contracts) is summarized in Table 10.1.

The uncertain demand is denoted by the random variable ξ whose support is $[0, +\infty]$. The probability density function (PDF) of ξ is $f(\cdot)$, cumulative distribution function (CDF) is $F(\cdot)$, and complementary CDF is $\bar{F}(\cdot)$. Assume that F is differentiable, increasing and $F(0) = 0$. We restrict our attention to demand distributions with an increasing failure rate (IFR). Let $z(\cdot) = f(\cdot)/\bar{F}(\cdot)$ be the failure rate. Then, $z(\xi_1) \leq z(\xi_2)$ for $0 \leq \xi_1 < \xi_2 < \infty$.

Our other modeling assumptions are the following:

A1: The bank, retailer and supplier are all risk neutral, and the market has no taxes, no transaction costs other than the bankruptcy costs).

A2: The market has full information (i.e., all the parameters in each model are common knowledge to the supplier, retailer, and bank).

A3: All bank loans are competitively priced (perfectly competitive banking sector), and the bank is assumed to face no bankruptcy risk.

A4: Both the supplier and retailer may face bankruptcy risk, depending on whether their liquid assets can cover their loan obligations.

A5: Both the supplier and retailer have long-term capital structure that is solely equity financed (assumed for convenience of exposition), and their objective is maximization of terminal (at the end of sales season) shareholders' wealth.

A6: The retailer and supplier are creditworthy and will repay their loan obligations (if any) to the extent possible.

From our assumption A5, we use retailer (supplier) to represent retailer's (supplier's) shareholders. From assumption A6, if the retailer (supplier) has enough wealth, he (she) will fully repay the loan principal and interest. Otherwise, he (she) will bankrupt and leave all available wealth, after bankruptcy costs (if any), to the creditors.

10.3 Bankrupt-Prone Supply Chains under Wholesale Price Contracts

The corporate finance literature traditionally argued on the separation of financing and operational decisions following the cornerstone work of Modigliani and Miller (1958). According to Modigliani and Miller's (M&M) theory, in a perfect market (i.e., no transaction or bankruptcy costs, no taxes, or agency considerations), the operational and financing decisions of a firm should be made independent of each other. However, the explicit consideration of bankruptcy costs in our research violates the premises of the M&M theory and leads to the interdependence of ordering and short-term financing decisions.

Xu and Birge (2004) are the first to incorporate bankruptcy costs within a newsvendor model, but assume that such costs are proportional to realized sales. Xu and Birge's focus was the study of interdependency of inventory decisions (quantity to order) and financing decisions (loan amount to borrow), and they point out the strong suboptimality of making such decisions independently for low-margin producers. Our work considers general bankruptcy costs, and contributes on understanding how the exact composition of the retailer's wealth (both working capital and collateral) affects the retailer's ordering behavior.

Our work in this section first studies the retailer's "bankrupt-prone newsvendor" problem and then analyzes the supplier's optimal wholesale price problem within a "selling to the bankrupt-prone newsvendor" framework explicitly accounting for the financial constraints of the retailer and with general bankruptcy costs. It clearly extends in nontrivial ways the work of Lariviere and Porteus (2001) analyzing the "selling to the newsvendor problem" as the Stackelberg game between the supplier and the retailer without any financial constraints. For more detail of the whole Section 10.3, please refer to Kouvelis and Zhao (2010).

TABLE 10.2 Time Line of Decisions

1 :	*Now:* The supplier (Stackelberg leader) presents a wholesale price only contract to the retailer.
2 :	*Now:* If the retailer accepts the contract, he determines an optimal order quantity and pays the supplier up front. If necessary, he borrows a loan from the bank.
3 :	*Now:* After receiving the loan request from the retailer, the bank decides an interest rate to make the loan competitively priced.
4 :	*The End of Sales Season:* The retailer repays the loan principal and associated interest. If his wealth is not enough to repay the loan, he declares bankruptcy.
5 :	*The End of Sales Season:* If the retailer declares bankruptcy, the bank gets involved into the bankruptcy process, and gets the remaining assets of the retailer after bankruptcy costs.

10.3.1 SEQUENCE OF EVENTS

The retailer orders from the supplier "now" at a given wholesale price w to satisfy the future uncertain demand. We assume that the bank, retailer, and supplier are risk neutral. All bank loans are competitively priced. The supplier (since $w \geq c$) and the bank are assumed to face no bankruptcy risk. However, the retailer might face bankruptcy risk, depending on whether or not his sales revenue and collateral can cover his loan obligations.

The notation is summarized in Table 10.1, and the assumptions is summarized in A1 to A6. Since in Section 10.3, we consider the bank financing, we use subscript b to denote bank financing.

The time line of our decisions is shown in Table 10.2.

10.3.2 THE NEWSVENDOR PROBLEM WITH BANKRUPTCY COSTS (THE BANKRUPT-PRONE NEWSVENDOR)

In this section, we consider the retailer's newsvendor problem. We will use r_b to denote the bank's interest rate to the retailer.

10.3.2.1 The Retailer's Profit Model. Let q_b be the retailer's order quantity for a given $0 < w(1 + r_f) \leq p$. Then, the loan amount he borrows now is $\max\{wq_b - y, 0\} = (wq_b - y)^+$. The required loan repayment at the end of sales season is $(wq_b - y)^+(1 + r_b(w, q_b))$, where $r_b(w, q_b)$ is the bank's interest rate. Note that there exists a unique $q_x(w)$ such that $(wq_x(w) - y)^+(1 + r_f) = (wq_x(w) - y)(1 + r_f) = x$. Let $\Omega = x + y(1 + r_f)$ be the retailer's total wealth in terms of working capital and collateral. Then, $q_x(w) = \frac{x+y(1+r_f)}{w(1+r_f)} = \frac{\Omega}{w(1+r_f)}$.

When the demand realized is low, it is better for the retailer to liquidate the collateral and use it to repay the loan and avoid a costly bankruptcy. As usual for such cases, the collateral can be liquidated by the retailer at no cost or loss of value (see Besanko and Thakor 1987, and Koziol 2007). Let $k_b(w, q_b)$ be the retailer's *bankruptcy threshold*, the minimal demand level that the retailer can repay the loan at the end of sales season. Then:

$$\begin{aligned} k_b(w, q_b) &= \frac{1}{p}\max\{(wq_b - y)^+(1 + r_b(w, q_b)) - x, 0\} \\ &= \frac{1}{p}[(wq_b - y)^+(1 + r_b(w, q_b)) - x]^+ \end{aligned} \tag{10.1}$$

From (10.1), $k_b(w, q_b)$ is a function of q_b and $r_b(w, q_b)$. Note that for any $q_b \leq q_x(w)$, $r_b(w, q_b) = r_f$ and $k_b(w, q_b) = 0$, since the loan can be fully secured by the collateral x. Moreover, (10.1) further implies that for a given $q_b \geq q_x(w)$, there is a one-to-one mapping between $k_b(w, q_b)$ and $r_b(w, q_b)$. We will see later in the paper that it is convenient to use q_b and $k_b(w, q_b)$, instead of q_b and $r_b(w, q_b)$, as the fundamental decision variables of the retailer and the bank. Then, once q_b and $k_b(w, q_b)$ are determined, we can easily find the interest rate $r_b(w, q_b)$ based on (10.1).

At the end of sales season, the retailer's sales is $p\min\{\xi, q_b\}$, and his collateral asset is x. Assume that $k_b(w, q_b) \leq q_b$. Note that when $q_b \geq q_x(w)$, and if the realized demand ξ is less than or equal to the bankruptcy threshold $k_b(w, q_b)$, then the retailer cannot repay the loan and has to declare bankruptcy. In this case, the retailer loses his sales and collateral (i.e., $p\xi + x$). As a result, the retailer's terminal cash flow is $0 = p\min\{\xi, q_b\} - p\min\{\xi, k_b(w, q_b)\}$, where the equality follows since $\xi \leq k_b(w, q_b) \leq q_b$. However, if the realized demand ξ is larger than $k_b(w, q_b)$, then the retailer can fully repay the loan, and his terminal cash flow is $p\min\{\xi, q_b\} + x - (wq_b - y)^+(1 + r_b(w, q_b)) = p\min\{\xi, q_b\} - pk_b(w, q_b) = p\min\{\xi, q_b\} - p\min\{\xi, k_b(w, q_b)\}$, where the first equality follows from (10.1) and the second one follows from $\xi > k_b(w, q_b)$. On the other hand, for $0 \leq q_b \leq q_x(w)$, the collateral x is enough to cover the loan obligation and thus the retailer's terminal cash flow is $p\min\{\xi, q_b\} + x + (y - wq_b)(1 + r_f) = p\min\{\xi, q_b\} - wq_b(1 + r_f) + \Omega$. Then, the retailer's expected terminal cash flow at the end of sales season, $\pi_b(w, q_b)$, is:

$$\pi_b(w, q_b) = \begin{cases} p\mathbb{E}(\min\{\xi, q_b\}) - p\mathbb{E}(\min\{\xi, k_b(w, q_b)\}), & \text{if } q_b \geq q_x(w), \quad (10.2a) \\ p\mathbb{E}(\min\{\xi, q_b\}) - wq_b(1 + r_f) + \Omega, & \text{if } 0 \leq q_b \leq q_x(w). \quad (10.2b) \end{cases}$$

Note that if $k_b(w, q_b) \geq q_b$, the retailer will bankrupt even if he sells all the inventory and the retailer's expected profit is non-positive, and thus the retailer has no incentive to borrow the bank loan. Therefore, in practice when the retailer borrows, we have $k_b(w, q_b) \leq q_b$, as shown in (10.2a).

10.3.2.2 Bank's Bankruptcy Costs and Retailer's Debt Capacity. A risk neutral bank offers a competitively priced loan to our retailer. For $q_b \geq q_x(w)$, the expected repayment the bank gets includes two parts: the repayment when the retailer gets into bankruptcy, $\Theta(\xi)$, which is the retailer's total wealth minus bankruptcy costs (to be modeled in detail below, see [10.5]), and the repayment when the retailer fully repays the loan obligation, $(wq_b - y)^+(1 + r_b(w, q_b)) =$

$pk_b(w, q_b) + x$ from (10.1). Then, the expected total repayment the bank gets from the retailer at the end of sales season is:

$$\Gamma(k_b(w, q_b)) = \int_0^{k_b(w,q_b)} \Theta(\xi)dF(\xi) + [pk_b(w, q_b) + x]\bar{F}(k_b(w, q_b)). \quad (10.3)$$

From (10.3), $\Gamma(k_b(w, q_x(w))) = \Gamma(0) = x$. For competitively priced loan and for $q_b \geq q_x(w)$ holds:

$$(wq_b - y)(1 + r_f) \equiv \Gamma(k_b(w, q_b)) \quad (10.4)$$

We refer to as (10.4) the *competitively priced loan equation*.

Failure of the retailer to repay the loan at the end of sales season leads to a costly bankruptcy. Our modeling of the bankruptcy process and related costs follows standard assumptions in the corporate finance literature (see Ang et al. 1982, Besanko and Thakor 1987, Tirole 2006, and Koziol 2007). The liquid assets of the retailer at the end of sales season include the realized sales proceeds and the liquidated value of his collateral. As a matter of process, these assets will be used first to cover all relevant fixed administrative costs B associated with declaring bankruptcy (i.e., fees to government agencies, lawyer and accountant fees for transactions required by the bankruptcy law). Banks engage in bankruptcy proceedings only if the available liquid assets of the retailer can cover the fixed administrative costs. In those cases, leftover assets after paying the administrative (fixed) bankruptcy cost are used by the bank in recouping some portion of the defaulted loan.

The corporate finance literature assumes that during bankruptcy a portion of the originally estimated value of the retailer's collateral is lost, and as part of the bankruptcy costs. The bank's urgency to sell the collateral assets via auctions and other liquidation markets, and bank's lack of deep expertise in the retailer's business, lead in losing a portion of $0 \leq \alpha_2 \leq 1$ of the collateral value. That is, for an originally estimated collateral value x, the bank realizes after bankruptcy liquidation a value of $(1 - \alpha_2)x$, with the $\alpha_2 x$ referred to as "depressed collateral bankruptcy costs." As in previous literature (Xu and Birge 2004 and Lai et al. 2009), bankruptcy costs include a portion of the sales proceeds of a bankrupt retailer. Let $0 \leq \alpha_1 \leq 1$ be the portion of realized sales revenue lost (i.e., not received by the bank), with $\alpha_1 p\xi$ referred to as "proportional to sales bankruptcy costs." In general, this $\alpha_1 p\xi$ can be due to the diversion of sales by the retailer, and the deceasing efforts to induce the demand when the retailer realizes that he has to declare bankruptcy. However, the usual explanation offered is that these are variable administrative costs in going through bankruptcy proceedings depending on the size of the bankrupt firm, and sales is a good approximate for firm size. Thus, in the case of bankruptcy, when $q_b \geq q_x(w)$, the repayment that the bank receives is:

$$\Theta(\xi) = [(1 - \alpha_1)p\xi + (1 - \alpha_2)x - B]^+ \quad (10.5)$$

when ξ is the demand realized, let

$$b = \frac{B - (1-\alpha_2)x}{(1-\alpha_1)p} \tag{10.6}$$

This is referred to as the "administrative bankruptcy threshold." We refer to the case of $b \le 0$ (i.e., $(1-\alpha_2)x \ge B$) as the "Collateral Protected Fixed Fee" case, and is the only case that the bank is guaranteed to receive some portion of its loan repaid in the case of bankruptcy. Otherwise, if $b \ge 0$, (i.e., $(1-\alpha_2)x \le B$), then only if the realized demand is higher than b, the bank is able to cover the fixed administrative cost B and justifies engaging in further bankruptcy proceedings. The bankruptcy case where $b \ge 0$ is referred to as the "NonProtected Fixed Fee" case, and in it the bank does not recoup any of its loan amount if the realized demand is less than b.

We can now state the bank's expected repayment from the loan transaction and its properties. By substituting (10.5) into (10.3) and using $B - (1-\alpha_2)x = (1-\alpha_1)pb$, for $q_b \ge q_x(w)$, we have that:

$$\Gamma(k_b(w,q_b)) = (1-\alpha_1)p\int_0^{k_b(w,q_b)}(\xi - b)^+ dF(\xi) + [pk_b(w,q_b) + x]\bar{F}(k_b(w,q_b)) \tag{10.7}$$

Let the partial derivative $\frac{\partial\Gamma(k_b(w,q_b))}{\partial k_b(w,q_b)} = p\bar{F}(k_b(w,q_b))G(k_b(w,q_b))$ where the function $G(\theta)$ for any real $\theta \ge 0$ is defined as:

$$G(\theta) = \begin{cases} 1 - \left[\alpha_1\theta + \frac{\alpha_2 x}{p} + \frac{B}{p}\right]z(\theta), & \text{if } \theta \ge \max\{b, 0\} \quad (10.8a) \\ 1 - \left[\theta + \frac{x}{p}\right]z(\theta), & \text{if } 0 \le \theta \le b \quad (10.8b) \end{cases}$$

We can verify that for a given $q_b \ge q_x(w)$, the bank's expected repayment function $\Gamma(k_b(w,q_b))$ is more concise if we use $k_b(w,q_b)$ as the bank's decision variable than using $r_b(w,q_b)$. Therefore, we use q_b as the retailer's decision variable to optimize his expected profit, $k_b(w,q_b)$ as the bank's decision variable to competitively price the bank loan, and w as the supplier's decision variable to optimize her expected profit. The bank can find $r_b(w,q_b)$ from (10.1) once $k_b(w,q_b)$ is determined.

In the case of no bankruptcy costs, $\alpha_1 = \alpha_2 x = B = 0$ so that $G(\theta) \equiv 1$ for any $\theta \ge 0$. However, we are interested in cases of some positive bankruptcy costs with $\alpha_1 + \alpha_2 x + B > 0$, and for such cases, $G(\theta) \le 1$. When $b \ge 0$, we have $\alpha_1\theta + \alpha_2\frac{x}{p} + \frac{B}{p} = \theta + \frac{x}{p}$ for $\theta = b$, following from the definition of b in (10.6). Then, $G(\theta)$ is continuous in θ, even if $b \ge 0$. Finally, $G(\theta)$ decreases in θ, since $z(\theta)$ increases in θ for IFR demand distributions.

Let's see what will happen if $G(0) \le 0$. If $b \ge 0$ (i.e., $B + \alpha_2 x \ge x$), then from (10.8b), $B + \alpha_2 x \ge x \ge \frac{p}{z(0)}$. Alternatively, if $b \le 0$ (i.e., $x \ge B + \alpha_2 x$), then from (10.8a), $x \ge B + \alpha_2 x \ge \frac{p}{z(0)}$. As a result, $\min\{B + \alpha_2 x, x\} \ge \frac{p}{z(0)}$. Since the bankruptcy costs $B + \alpha_2 x$ are so large, the bank will never issue the retailer a loan that cannot be secured by his collateral assets x. Therefore, $q_{\max}(w) = q_x(w)$. Furthermore, we can show that in this case, the retailer's total wealth, $\Omega = x + y(1 + r_f)$, is enough for him to order the traditional newsvendor quantity for any $0 < w \leq \frac{p}{1+r_f}$, which we are not interested in. Consequently, we assume $G(0) > 0$ from now on to the end of Section 10.3.

Lemma 10.1. *For* $0 \leq q_b \leq q_x(w)$, $k_b(w, q_b) = 0$ *and* $r_b(w, q_b) = r_f$. *For* $q_b \geq q_x(w)$, *we have two cases, based on the value of* $G(k_b(w, q_b))$*:*

1. $G(k_b(w, q_b)) \geq 0$ *for* $k_b(w, q_b) \geq 0$. *Then,* $q_b \leq q_{\max}(w) < +\infty$ *where*

$$q_{\max}(w) = \begin{cases} q_x(w) + \frac{p\int_0^{+\infty} \bar{F}(\xi)d\xi - \alpha_2 x - B}{w(1+r_f)}, & \text{if } b \leq 0 \quad (10.9a) \\ q_x(w) + \frac{p\int_b^{+\infty} \bar{F}(\xi)d\xi - x}{w(1+r_f)}, & \text{if } b \geq 0 \quad (10.9b) \end{cases}$$

Let $k_{\max}(w) = r_{\max}(w) = +\infty$. *Then, as* $q_b \to q_{\max}(w)$, $k_b(w, q_b) \to k_{\max}(w)$ *and* $r_b(w, q_b) \to r_{\max}(w)$.

2. $G(0) > 0$ *and* $G(+\infty) < 0$. *There exists a unique* $0 < k_{\max}(w) < +\infty$ *so that* $G(k_{\max}(w)) = 0$, *a unique* $q_{\max}(w) \geq q_x(w)$ *so that* $k_b(w, q_{\max}(w)) = k_{\max}(w)$, *and a unique* $r_{\max}(w) \geq r_f$ *so that* $r_b(w, q_{\max}(w)) = r_{\max}(w)$. *Then, we have* $q_b \leq q_{\max}(w)$, $k_b(w, q_b) \leq k_{\max}(w)$, *and* $r_b(w, q_b) \leq r_{\max}(w)$.

From Lemma 10.1, we see that for the two cases, $q_b \leq q_{max} < +\infty$. That is, the retailer's order quantity is finite, even if the bank does not place a credit line limit on the loan. This $q_{\max}(w)$ is referred as the "finite debt capacity" of the retailer. Xu and Birge (2004) establish a similar result but only for the case of proportional to sales bankruptcy costs and for $0 < \alpha_1 \leq 1$. Our result on the existence of the finite debt capacity is in a more general context of the bankruptcy costs.

In the first case of Lemma 10.1, $G(+\infty) \geq 0$ so that $\alpha_1 = 0$ (i.e., no bankruptcy cost proportional to the realized sales). Let us focus on the case (10.9a) (i.e., $b \leq 0$). Note that $\int_0^{+\infty} \bar{F}(\xi)d\xi = \int_0^{+\infty} \xi f(\xi)d\xi = \mu$, where μ is the finite mean of the uncertain demand. From (10.9a), $q_{\max}(w) = q_x(w) + \frac{p\mu - \alpha_2 x - B}{w(1+r_f)} = \frac{p\mu + (1-\alpha_2)x + y(1+r_f) - B}{w(1+r_f)}$ (i.e., $wq_{\max}(w)(1 + r_f) = p\mu + (1 - \alpha_2)x + y(1 + r_f) - B$). Note that $p\mu$ is the largest expected sales from the realized demand and $p\mu + (1 - \alpha_2)x + y(1 + r_f) - B$ is the largest expected total wealth (after adjusting for the bankruptcy costs). Then, if $q_b \geq q_{\max}(w)$, the retailer's ordering cost will exceed the largest expected total wealth, and he will get a negative expected profit. On the other hand, from the bank's perspective, $(wq_{\max}(w) - y)(1 + r_f) = p\mu + (1 - \alpha_2)x - B$, where the right hand side is $\Gamma(+\infty)$ from (10.7). Thus, if the retailer orders $q_{\max}(w)$, the bank has to charge an interest rate so that the bankruptcy threshold is infinite, otherwise the bank cannot break even. However, in this case, as the retailer will bankrupt with certainty, he will not borrow.

For the second case of Lemma 10.1, we can see for each $q_x(w) \leq q_b \leq q_{\max}(w)$, there can be two values of $r_b(w, q_b)$ and two values of $k_b(w, q_b)$ that

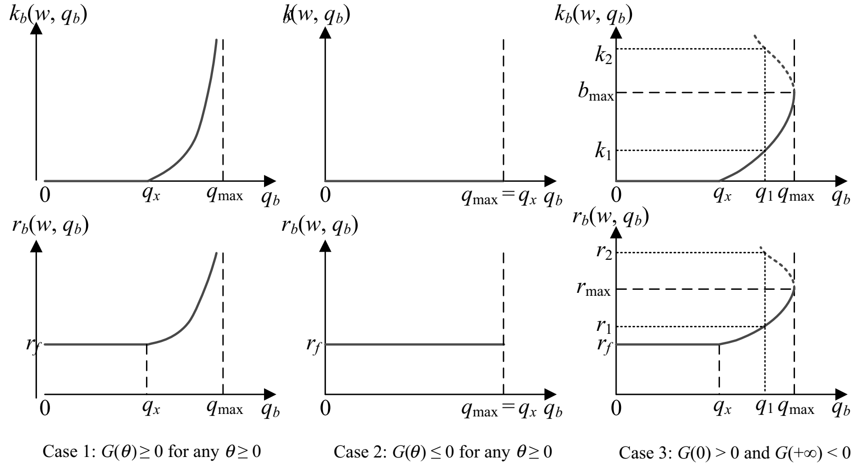


FIGURE 10.1 **Relationships between $k_b(w, q_b)$ and q_b and between $r_b(w, q_b)$ and q_b.**

end up with the bank loan being competitively priced. For example, k_2 (k_1) is the corresponding bankruptcy threshold if the bank offers r_2 (r_1) as the interest rate to the retailer when he orders the quantity q_1. Please refer to Case 3 in Figure 10.1. From (10.2a), note that the retailer's profit is smaller if the bank offers r_2 (instead of r_1) as the interest rate of the bank loan. In a highly competitive financial market, the retailer will choose another bank to get the loan. Therefore, eventually, the equilibrium solution between the bank and retailer will end up with an interest rate $r \leq r_{max}(w)$, and bankruptcy threshold $k_b(w, q_b) \leq k_{max}(w)$, as shown in Figure 10.1.

The relationships between q_b and $k_b(w, q_b)$ and between q_b and $r_b(w, q_b)$ are summarized as follows.

LEMMA 10.2. *For $q_x(w) \leq q_b \leq q_{max}(w)$, we have $G(k_b(w, q_b)) \geq 0$. Thus, $\Gamma(k_b(w, q_b))$ increases in $k_b(w, q_b)$. Also, $k_b(w, q_b)$ is increasing and convex in q_b. Finally, $r_b(w, q_b)$ is increasing in q_b.*

From Lemma 10.2, $k_b(w, q_b)$ and $r_b(w, q_b)$ increase in q_b. From (10.1), the bankruptcy threshold $k_b(w, q_b)$ and the loan principal plus interest have a linear relationship. Then, the loan principal plus interest is also increasing and convex in the order quantity. However, although we know that the interest rate always increases in the order quantity, we do not know whether it is convex in it or not.

So far, we explain that for a given wholesale price w, if the retailer orders a quantity larger than $q_{max}(w)$, then the competitively priced loan equation cannot be met. Our detailed analysis indicates that for any given $0 < w \leq \frac{p}{1+r_f}$, we have $\bar{F}^{-1}\left(\frac{w(1+r_f)}{p}\right) \leq q_{max}(w)$ (i.e., the traditional newsvendor order quantity is within the retailer's debt capacity). In Section 10.3.3.1, we establish that for any retailer, his optimal order quantity is no larger than the traditional newsvendor order quantity. As a result, when the retailer orders the optimal quantity, the competitively priced loan equation can always be met. Consequently, we do not further discuss the retailer's debt capacity from now on to the end of Section 10.3.

10.3.2.3 Retailer's Optimal Order Quantity for a Fixed Wholesale Price. We now determine the retailer's optimal order quantity. By analyzing the first order derivative, we show that the retailer's profit $\pi_b(w, q_b)$ as defined in (10.2a) and (10.2b) are quasi-concave in q_b. The following Proposition 10.1 gives the optimal order quantity of the retailer, denoted as $q_b^*(w)$. The corresponding bankruptcy threshold is denoted as $k_b^*(w) = k_b(w, q_b^*(w))$.

Proposition 10.1 *The retailer's profit $\pi_b(w, q_b)$ is quasi-concave in q_b. The optimal order quantity depends on the retailer's wealth and is specified according to the following cases:*

1. *When $\Omega \geq \Omega_1(w)$, where $\Omega_1(w) = w(1 + r_f)\bar{F}^{-1}\left(\frac{w(1+r_f)}{p}\right)$, the optimal order quantity $q_b^*(w) = \bar{F}^{-1}\left(\frac{w(1+r_f)}{p}\right)$, (i.e., the traditional newsvendor solution). Also, $k_b^*(w) = 0$.*
2. *For $\Omega_2(w) \leq \Omega \leq \Omega_1(w)$, where $\Omega_2(w) = w(1 + r_f)\bar{F}^{-1}\left(\frac{w(1+r_f)}{pG(0)}\right)$ if $pG(0) > w(1 + r_f)$ and $\Omega_2(w) = 0$ otherwise, the optimal order quantity $q_b^*(w) = q_x(w) = \frac{\Omega}{w(1+r_f)}$. Also, $k_b^*(w) = 0$.*
3. *For $\Omega_3(w) \leq \Omega \leq \Omega_2(w)$, where $\Omega_3(w) = w(1 + r_f)\bar{F}^{-1}\left(\frac{w(1+r_f)}{pG(\max\{b,0\})}\right) - (p \max\{b, 0\} + x)\bar{F}(\max\{b, 0\}) + x$ if $pG(\max\{b, 0\}) > w(1 + r_f)$ and $\Omega_3(w) = 0$ otherwise, the optimal order quantity $q_b^*(w)$ is solved from $p\bar{F}(q_b) - \frac{w(1+r_f)}{G(k_b(w,q_b))} = 0$ where $G(k_b(w, q_b))$ is defined by (10.8b). Also, $0 \leq k_b^*(w) \leq \max\{b, 0\}$.*
4. *Finally, for $\Omega \leq \Omega_3(w)$, the optimal order quantity $q_b^*(w)$ is solved from $p\bar{F}(q_b) - \frac{w(1+r_f)}{G(k_b(w,q_b))} = 0$ where $G(k_b(w, q_b))$ is defined by (10.8a). Also, $k_b^*(w) \geq \max\{b, 0\} \geq 0$.*

From Proposition 10.1, in the presence of bankruptcy risks and costs, the retailer's ordering decision is not just determined from the demand distribution with the critical ratio, as in the traditional newsvendor formulas, but depends also on the exact composition of his wealth in terms of working capital and collateral. The wealthier the retailer is, the harder the retailer gets into bankruptcy. Therefore, if the retailer's wealth $\Omega \geq \Omega_1(w)$, the retailer's wealth is enough for him to order the newsvendor quantity. Then, the retailer behaves exactly as a traditional newsvendor, and does not have bankruptcy risks. This is Case 1 of Proposition 10.1. When the retailer's wealth decreases, it may not be enough anymore to purchase the newsvendor amount, and as the retailer does not want to lose his current wealth by getting a loan and incurring bankruptcy risks, thus he orders what his wealth allows him. This is Case 2 of Proposition 10.1. When the retailer's wealth decreases further, his order quantity by ordering with all his wealth and without borrowing a bank loan also decreases, which is no longer the optimal order quantity to the retailer. Instead, he prefers to borrow a bank loan and orders a quantity larger than the amount affordable by his total wealth. Since in this case his loan amount is not too large, his bankruptcy risk is small with a bankruptcy threshold between 0 and $\max\{b, 0\}$. This is Case 3 of Proposition 10.1. Finally, if the retailer's wealth becomes even smaller, the retailer has to increase his loan amount, and his bankruptcy threshold becomes larger than $\max\{b, 0\}$. This is Case 4 of Proposition 10.1.

Here we want to mention that $\Omega_1(w)$ is only a function of w but not the retailer's collateral x and fixed administrative cost B. However, both $\Omega_2(w)$ and $\Omega_3(w)$ are functions of w, x, and B, since $G(0)$ and $G(b)$ are functions of x and B. Just for ease of exposition, we suppress x and B from the definitions of $\Omega_2(w)$ and $\Omega_3(w)$. Based on the Properties of $\Omega_1(w)$, $\Omega_2(w)$, and $\Omega_3(w)$, we can plot the three curves, $\Omega = \Omega_1(w)$, $\Omega = \Omega_2(w)$ and $\Omega = \Omega_3(w)$ in the x-y space, and this way partitions the x-y space into four regions, as shown in Figure 10.2. Note that $\Omega = \Omega_1(w)$ is a straight line in x-y space, but $\Omega = \Omega_2(w)$ and $\Omega = \Omega_3(w)$ are not straight lines.

We refer to the region in which $q_b^*(w)$ satisfies $p\bar{F}(q_b) - \frac{w(1+r_f)}{G(k_b(w,q_b))} = 0$ and $G(k_b(w, q_b))$ is as defined in (10.8a) as *Bankruptcy Region A*; the region in which $q_b^*(w)$ satisfies $p\bar{F}(q_b) - \frac{w(1+r_f)}{G(k_b(w,q_b))} = 0$ where $b \geq 0$ and $G(k_b(w, q_b))$ is as defined in (10.8b) as *Bankruptcy Region B*; the region in which $q_b^*(w) = q_x(w)$ as *Ordering with All Wealth Region*; and finally the region where $q_b^*(w) = \bar{F}^{-1}\left(\frac{w(1+r_f)}{p}\right)$ as *Traditional Newsvendor Region*. Notice that in Bankruptcy Region A, $k_b^*(w) \geq \max\{b, 0\}$ and in Bankruptcy Region B, $0 \leq k_b^*(w) \leq \max\{b, 0\}$. Therefore, if $b < 0$, then in Bankruptcy Region A, $k_b^*(w) \geq 0$, and Bankruptcy Region B does not exist. Moreover, for the case that $G(0) \leq 0$, the retailer is in the Traditional Newsvendor Region for any $0 < w \leq \frac{p}{1+r_f}$.

For the general situation, those four regions are clearly identified for both a high administrative bankruptcy cost case (Case 1, with $B \geq (1 - \alpha_2)x_a(w)$) and a low one (Case 2, with $B < (1 - \alpha_2)x_a(w)$), where $x_a(w)$ solves $x = \Omega_2(w)$ (recall that $\Omega_2(w)$ is a function of x).

From Figure 10.2, we observe that increases in wealth (either working capital, collateral, or both) lead to increases in order quantities as we successively traverse the x-y space from Bankruptcy Region A, to Bankruptcy Region B, Ordering with All Wealth Region, and finally Traditional Newsvendor Region. For example, for a fixed collateral value x_0 (x_0 is selected in Bankruptcy Region A), increases in working capital (from y_0 to y_1 to y_2 to y_3) lead to successively increasing optimal order quantity (the sensitivity of order quantity to retailer's wealth is formally established in Section 10.3.2.4). Similar traversal of all four ordering regions is possible through keeping a fixed working capital level y_0 and successively increasing x. Finally, diagonal moves involving changes in both x and y from the lower left corner of the graph to the right upper corner may also lead to traversal of all four ordering regions.

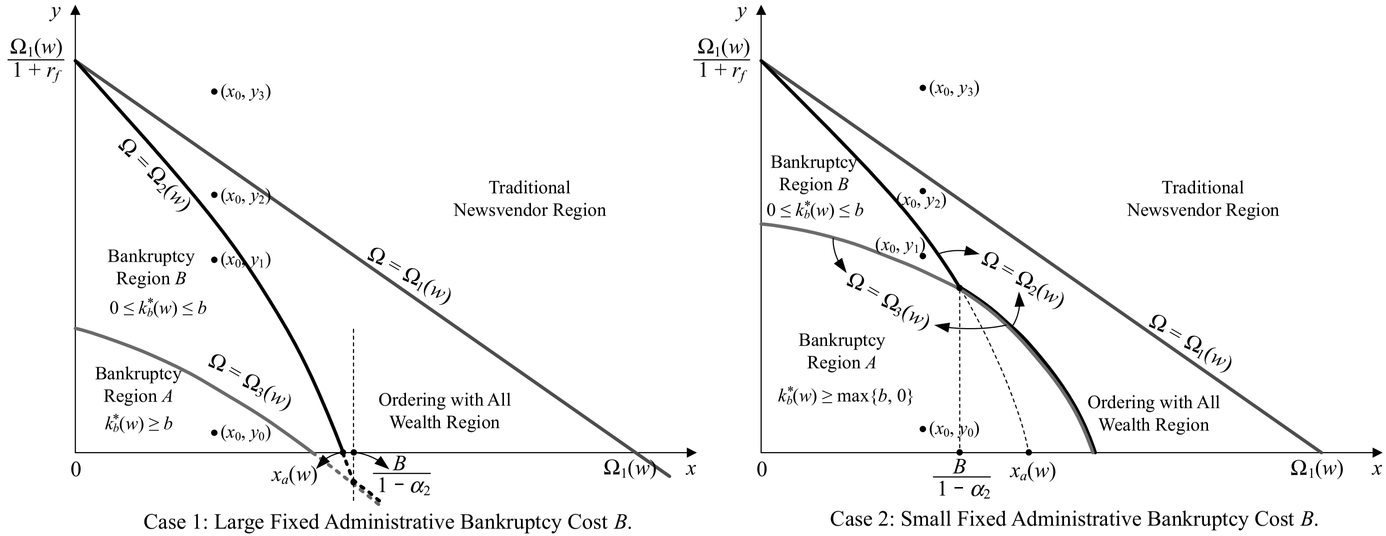


FIGURE 10.2 **The retailer's optimal solution depending on wealth. In case 1, B is large: $B \geq (1 - \alpha_2)x_a(w)$. In case 2, B is small: $B < (1 - \alpha_2)x_a(w)$.**

10.3.2.4 Sensitivity Analysis. In this section, we conduct sensitivity analysis on x and y in an effort to answer the question on the optimal mix of working capital and collateral the retailer should use to borrow from the bank. Notice that the bank charges a premium on the interest rate because of the bankruptcy costs. For y, intuitively, he should use all of it because for each dollar he uses, his benefit, $1 + r_b(w, q_b)$, is larger than or equal to the cost, $1 + r_f$. Proposition 10.2 justifies this result. Let $\pi_b^*(w) = \pi_b(w, q_b^*(w))$.

Proposition 10.2 *In Bankruptcy Regions A and B, $q_b^*(w)$ and $\pi_b^*(w)$ increase in y and $k_b^*(w)$ decreases in y, and $\frac{dq_b^*(w)}{dy} \leq \frac{1}{w}$ and $\frac{d\pi_b^*(w)}{dy} = (1 + r_f)\left[\frac{p\bar{F}(q_b^*(w))}{w(1+r_f)} - 1\right]$. In the Ordering with All Wealth Region, $q_b^*(w)$ and $\pi_b^*(w)$ increase in y, and $\frac{dq_b^*(w)}{dy} = \frac{1}{w}$ and $\frac{d\pi_b^*(w)}{dy} = (1 + r_f)\left[\frac{p\bar{F}(q_b^*(w))}{w(1+r_f)} - 1\right]$. Finally, in the Traditional Newsvendor Region, both $q_b^*(w)$ and $\pi_b^*(w)$ do not change in y.*

Next, we consider the effect of the collateral x. There are two functions of x: First, it decreases the bankruptcy threshold $k_b(w, q_b)$ for a fixed order quantity q_b, and thus helps the retailer get a lower interest rate for the bank loan; and second, it increases the retailer's order quantity q_b. However, the increase in order quantity may affect the bankruptcy threshold. This interaction effect between order quantity and bankruptcy threshold complicates the effect of increased x on the retailer's order quantity and profit. The formal result for reasonable cases of collateral use (i.e., $\alpha_1 \geq \alpha_2$) follows in Proposition 10.3.

Proposition 10.3 *In Bankruptcy Regions A and B, $q_b^*(w)$ and $\pi_b^*(w)$ increase in x and $k_b^*(w)$ decreases in x, and $\frac{dq_b^*(w)}{dx} \leq \frac{1}{w(1+r_f)}$ and $\frac{d\pi_b^*(w)}{dx} \leq \frac{p\bar{F}(q_b^*(w))}{w(1+r_f)} - 1$. In the Ordering with All Wealth Region, both $q_b^*(w)$ and $\pi_b^*(w)$ increase in x, and $\frac{dq_b^*(w)}{dx} = \frac{1}{w(1+r_f)}$ and $\frac{d\pi_b^*(w)}{dx} = \frac{p\bar{F}(q_b^*(w))}{w(1+r_f)} - 1$. Finally, in the Traditional Newsvendor Region, both $q_b^*(w)$ and $\pi_b^*(w)$ do not change in x.*

With the above results in place, we can now state the formal result on the effectiveness of collateral and working capital in cases the retailer needs to borrow.

Proposition 10.4 *In the Ordering with All Wealth and Traditional Newsvendor Regions, the working capital and collateral have the same effects on the retailer's optimal order quantity and profit. However, in the Bankruptcy Region(s), the collateral is less effective than the working capital in increasing his order quantity and profit if $\alpha_2 > 0$ or $B > 0$, and has the same effects as the working capital otherwise (i.e., $\alpha_2 = B = 0$).*

In the Ordering with All Wealth Region and the Traditional Newsvendor Region, the working capital and collateral have the same effects on retailer's optimal order quantity and profit, since the retailer can fully utilize them. However, in the Bankruptcy Region(s), since the collateral loses some value when $\alpha_2 > 0$, the working capital is more effective for the retailer. Even if $\alpha_2 = 0$ but $B > 0$, portion of the collateral has to be used to pay the administrative bankruptcy cost B first, and thus the working capital continues to be more effective.

10.3.3 SELLING TO THE BANKRUPT-PRONE NEWSVENDOR

In this section, we solve the supplier's problem in setting the optimal wholesale price in order to maximize her profit. In Section 10.3.3.1, we first study how the retailer's optimal order quantity changes in the supplier's wholesale price, and show that there is a 1-to-1 mapping between w and $q_b^*(w)$.

10.3.3.1 The Optimal Order Quantity for Fixed Retailer's Wealth.

In this section, we discuss how the retailer's optimal order quantity changes in the wholesale price. For a given w, from Proposition 10.1, we can get the optimal order quantity $q_b^*(w)$ by comparing the magnitude of Ω, $\Omega_1(w)$, $\Omega_2(w)$, and $\Omega_3(w)$. However, $\Omega_1(w)$, $\Omega_2(w)$, and $\Omega_3(w)$ are all functions of w. To understand how the optimal order quantity is influenced by the wholesale price, we need to investigate how $\Omega_1(w)$, $\Omega_2(w)$, and $\Omega_3(w)$ change in w, as shown in the following lemma.

Lemma 10.3. *We have $\Omega_1(w)$, $\Omega_2(w)$ for $G(0) > 0$, and $\Omega_3(w)$ for $G(\max\{b, 0\}) > 0$ are concave in w. Let $\bar{\Omega} = \max_{0<w(1+r_f)\le p}\{\Omega_1(w)\}$, $\underline{\Omega} = \max_{0<w(1+r_f)\le pG(0)}\{\Omega_2(w)|G(0) > 0\}$ and $\underline{\Omega} = 0$ otherwise, and $\underline{\underline{\Omega}} = \max_{0<w(1+r_f)\le pG(\max\{b,0\})}\{\Omega_3(w)|G(\max\{b, 0\}) > 0\}$ and $\underline{\underline{\Omega}} = 0$ otherwise. Then, $\underline{\underline{\Omega}} \le \underline{\Omega} \le \bar{\Omega}$. Also,*

1. *For $\Omega \le \bar{\Omega}$, there exist at most two values of w, $0 < w^1 \le w^2 \le \frac{p}{1+r_f}$, so that $\Omega_1(w) = \Omega$;*
2. *For $\Omega \le \underline{\Omega}$, there exist at most two values of w, $w^1 \le w^3 \le w^4 \le w^2$, so that $\Omega_2(w) = \Omega$;*
3. *For $\Omega \le \underline{\underline{\Omega}}$, there exist at most two values of w, $w^3 \le w^5 \le w^6 \le w^4$, so that $\Omega_3(w) = \Omega$.*

Based on Lemma 10.3, when $\Omega \ge \bar{\Omega}$, the retailer is in the Traditional Newsvendor Region and orders exactly as the traditional newsvendor model. When $\underline{\Omega} \le \Omega \le \bar{\Omega}$, w^1 and w^2 exist but not w^3, w^4, w^5, and w^6, and the retailer orders $q_x(w)$ using all his wealth for $w^1 \le w \le w^2$, and follows the traditional newsvendor for $w \le w^1$ and $w \ge w^2$. When $\underline{\underline{\Omega}} \le \Omega \le \underline{\Omega}$, w^1, w^2, w^3, and w^4 all exist, but not w^5 and w^6. In this case, the retailer follows the traditional newsvendor for $w \le w^1$ and $w \ge w^2$, orders using all his wealth for $w^1 \le w \le w^3$ and $w^4 \le w \le w^2$, and is in Bankruptcy Region A for $w^3 \le w \le w^4$. Finally, when $\Omega \le \underline{\underline{\Omega}}$, w^1 through w^6 all exist. Similarly, we can analyze the different regions corresponding to the ranges of w. Please refer to Figure 10.3.

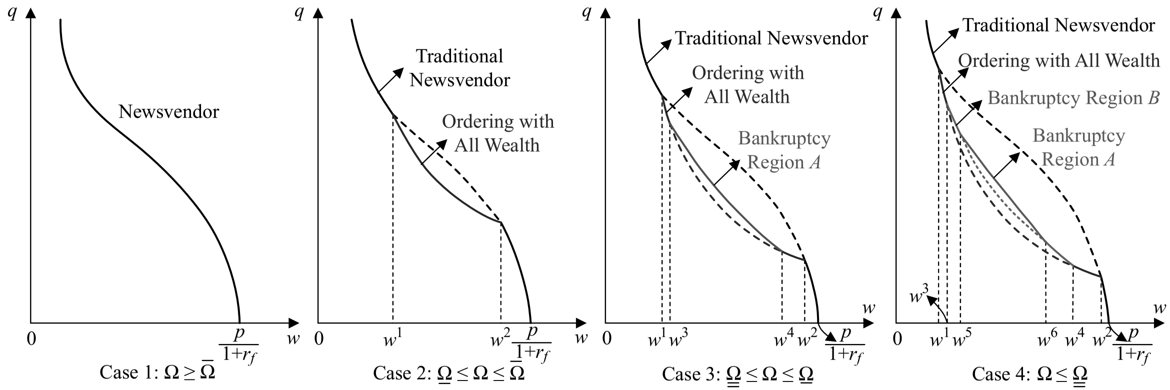


FIGURE 10.3 **Retailer's Response Curves $q_b^*(w)$.**

Proposition 10.5 *The optimal order quantity $q_b^*(w)$ is continuous and decreasing in the wholesale price w. Moreover, the relationship of the retailer's optimal order quantity and wholesale price is:*

1. *If $\Omega \geq \bar{\Omega}$, then the optimal order quantity is always $q_b^*(w) = \bar{F}^{-1}\left(\frac{w(1+r_f)}{p}\right)$, and if $\Omega \leq \bar{\Omega}$, then the optimal solution is $q_b^*(w) = \bar{F}^{-1}\left(\frac{w(1+r_f)}{p}\right)$ for $0 < w \leq w^1$ and $w^2 \leq w \leq \frac{p}{1+r_f}$.*
2. *If $\underline{\Omega} \leq \Omega \leq \bar{\Omega}$, then the optimal solution $q_b^*(w)$ is $q_x(w)$ for $w^1 \leq w \leq w^2$, and if $\Omega \leq \underline{\Omega}$, then the optimal solution is $q_x(w)$ for $w^1 \leq w \leq w^3$ and $w^4 \leq w \leq w^2$.*
3. *If $\underline{\underline{\Omega}} \leq \Omega \leq \underline{\Omega}$, then the optimal solution $q_b^*(w)$ is in the Bankruptcy Region A for $w^3 \leq w \leq w^4$. If $\Omega \leq \underline{\underline{\Omega}}$, then it is in the Bankruptcy Region B for $w^3 \leq w \leq w^5$ and $w^6 \leq w \leq w^4$, and is in the Bankruptcy Region A for $w^5 \leq w \leq w^6$.*

For a fixed retailer's wealth, we can easily show that in the Ordering with All Wealth Region, due to limited wealth constraining the placed order, the retailer orders a smaller quantity than the traditional newsvendor order quantity. However, in the Bankruptcy Region B, since the order quantity with all wealth is too small, the retailer decides to order a larger quantity by borrowing and incurring a small bankruptcy risk with $0 \leq k_b^*(w) \leq \max\{b, 0\}$. But, this order quantity is still less than the traditional newsvendor order quantity. Finally, in the Bankruptcy Region A, the order quantity with small bankruptcy risk (i.e., $0 \leq k_b^*(w) \leq \max\{b, 0\}$) is no longer optimal, and the retailer wants a larger order quantity with a larger bankruptcy risk $k_b^*(w) \geq \max\{b, 0\}$. Still, this order quantity is smaller than the traditional newsvendor order quantity. Please refer to the four cases of Figure 10.3.

10.3.3.2 The Stackelberg Game Between the Supplier and Retailer. In the game between the supplier (she) and retailer (he), the supplier is the Stackelberg leader and the retailer is the follower. Since it is a wholesale price contract, the optimal wholesale price will be set larger than the production cost, and thus the supplier does not face any bankruptcy risks.

Without loss of generality, and for presentation convenience, we present our detailed analysis of the selling to the bankrupt-prone newsvendor problem for the Collateral Protected Fixed Fee case, (i.e., $x \geq \frac{B}{1-\alpha_2}$) or alternatively, $b \leq 0$. The other case (i.e., $x \leq \frac{B}{1-\alpha_2}$ or $b \geq 0$) is analyzed in a similar fashion. For our case of interest, we can work with a simplified version of Figure 10.3 (see Figure 10.4) since $\underline{\underline{\Omega}} = \underline{\Omega}$ and thus $w^3 = w^5$ and $w^4 = w^6$. Now, Bankruptcy Regions A and B shrink into one region, Bankruptcy Region A, and for ease of exposition, we refer to it simply as *Bankruptcy Region*. In this region, the function $G(\theta)$ is given by (10.8a).

The retailer's response curve is $q_b^*(w)$. Proposition 10.5 establishes the 1-to-1 relationship between w and q_b^*. Let $w_b^*(q_b)$ be the inverse function of $q_b^*(w)$, i.e.,

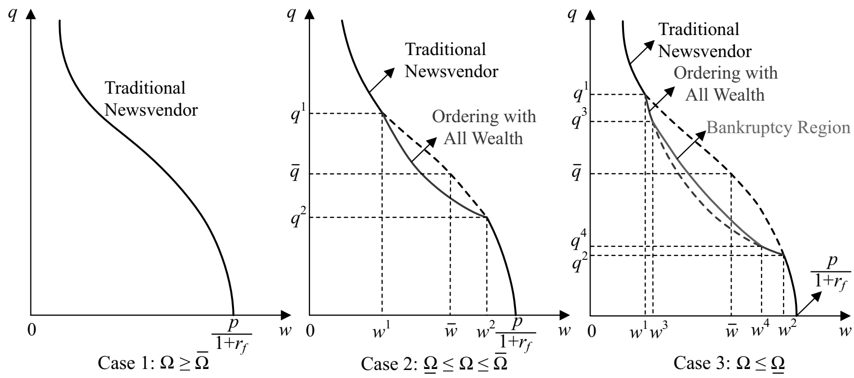


FIGURE 10.4 **The retailer's response curves $q_b^*(w)$, If $x \geq \frac{B}{1-\alpha_2}$.**

for any $q_b \geq 0$, if the supplier chooses $w_b^*(q_b)$ as her wholesale price, then the q_b will be the optimal order quantity of the retailer. Then, the supplier's profit at the end of sales season is given by:

$$\Pi_b^*(q_b) = (w_b^*(q_b)q_b - cq_b)(1 + r_f)$$

We use q_b^{**} and w_b^{**} to denote the *global* Stackelberg equilibrium retailer's order quantity and supplier's wholesale price. Note that w_b^{**} is the optimal wholesale price to the supplier.

Based on Figure 10.4, we have three cases: $\Omega \geq \bar{\Omega}$, $\underline{\Omega} \leq \Omega \leq \bar{\Omega}$, and $\Omega \leq \underline{\Omega}$. For the first two cases, the retailer does not have bankruptcy risk. Since the retailer will either order the traditional newsvendor quantity or order with all wealth, the supplier's problem is straightforward to solve. As a result, we focus on the third case, i.e., $\Omega \leq \underline{\Omega}$, in this section. Note that the retailer has bankruptcy risks in this case.

Let q^i be the retailer's optimal order quantity corresponding to w^i, for $i = 1, 2, 3$, and 4 (please see Case 3 in Figure 10.4). In the Bankruptcy Region where $q^4 \leq q_b \leq q^3$, the supplier's problem can be stated explicitly as follows:

$$\Pi_b^*(q_b) = \max_{q^4 \leq q_b \leq q^3} (w_b^*(q_b)q_b - cq_b)(1 + r_f)$$

$$\text{sub. to} \quad p\bar{F}(q_b)G(k_b(w_b^*(q_b), q_b)) - w(1 + r_f) = 0, \tag{10.10}$$

$$(w_b^*(q_b)q_b - y)(1 + r_f) - \Gamma(k_b(w_b^*(q_b), q_b)) = 0, \tag{10.11}$$

where $\Gamma(k_b(w_b^*(q_b), q_b))$ is defined in (10.3). Equation (10.10) is the retailer's first order condition to determine the optimal order quantity for a given wholesale price, and (10.11) is the bank's competitively priced loan equation to determine the bankruptcy threshold $k_b(w_b^*(q_b), q_b)$ (and then the interest rate from the bankruptcy threshold $k_b(w_b^*(q_b), q_b)$).

Proposition 10.6 shows that there is a unique equilibrium solution in the Bankruptcy Region.

Proposition 10.6 *Let the failure rate of the demand distribution, $z(\xi)$, be both increasing and convex on its support. In the Bankruptcy Region, the supplier's expected profit, $\Pi_b^*(q_b)$, under the constraints (10.10) and (10.11), is quasi-concave in q_b. Then,*

1. *The equilibrium wholesale price and order quantity in the Bankruptcy Region, denoted by $\hat{w}_b^*$ and $\hat{q}_b^*$, respectively, are unique and can be solved from the system of (10.10), (10.11), and*

$$\bar{F}(k_b(w_b^*(q_b), q_b))G(k_b(w_b^*(q_b), q_b))[w_b^*(q_b)(1 - q_b z(q_b)) - c] + cq_b\bar{F}(q_b)\frac{\partial G(k_b(w_b^*(q_b), q_b))}{\partial k_b(w_b^*(q_b), q_b)} = 0 \quad (10.12)$$

where the function $G(k_b(w_b^*(q_b), q_b))$ is defined in (10.8a).

2. *$\hat{q}_b^* \leq \bar{q}$ where $\bar{q}$ satisfies $\bar{q}z(\bar{q}) = 1$.*

The class of demand distributions that have increasing and convex failure rates includes uniform, exponential, power and Weibull with the shape parameter greater than or equal to 2, truncated normal distributions, and so on. Please refer to Kouvelis and Zhao (2008) for more details.

From Proposition 10.6, when $\alpha_1 = \alpha_2 x = B = 0$ (i.e., there are no bankruptcy costs) $G(k_b(w, q_b)) \equiv 1$ and (10.10), (10.11), and (10.12) are reduced to

$$p\bar{F}(q_b) - w_b^*(q_b)(1 + r_f) = 0 \quad (10.13)$$

$$(w_b^*(q_b)q_b - y)(1 + r_f) - \left[p\int_0^{k_b(w_b^*(q_b), q_b)} \bar{F}(\xi)d\xi + x\right] = 0 \quad (10.14)$$

$$p\bar{F}(q_b)(1 - q_b z(q_b)) - c(1 + r_f) = 0 \quad (10.15)$$

Equation (10.15) is obtained from (10.12) by substituting $w_b^*(q_b)(1 + r_f) = p\bar{F}(q_b)$ and $\frac{\partial G(k_b(w_b^*(q_b), q_b))}{\partial k_b(w_b^*(q_b), q_b)} = 0$. Note that (10.13) is the traditional newsvendor response curve, which is independent of the interest rate of the bank loan. Also, (10.15) is the result established by Lariviere and Porteus (2001). Note that equation (10.15) is independent of the supplier's wholesale price and the interest rate of the bank loan. Then, we let $\bar{q}_s^*$ be the unique solution of (10.15) and let $\bar{w}_s^* = \frac{p\bar{F}(\bar{q}_s^*)}{1+r_f}$. That is, $\hat{q}_b^* = \bar{q}_s^*$. Once we have $\hat{q}_b^*$, we determine $\hat{w}_b^* = w_b^*(\hat{q}_b^*) = \bar{w}_s^*$ from (10.13), and finally, determine $k_b(\hat{w}_b^*, \hat{q}_b^*)$ from (10.14).

Next, we focus on the general case with bankruptcy costs. Let q^4 be the order quantity corresponding to w^4, and q^2 corresponding to w^2. Notice that q^4 is the lower bound of the Bankruptcy Region so that $\hat{q}_b^*$ is feasible only if $\hat{q}_b^* \geq q^4$.

Lemma 10.4. *We have that $\hat{q}_b^* \leq \bar{q}_s^*$. That is, the equilibrium order quantity in the Bankruptcy Region is less than or equal to the traditional newsvendor equilibrium order quantity. If $c = 0$, then $\hat{q}_b^* = \bar{q}_s^* = \tilde{q}$, $\hat{w}_b^* \leq \bar{w}_s^*$, and $\Pi_b^*(\hat{q}_b^*) \leq \Pi_b^*(\bar{q}_s^*)$. In other words, when c is small and can be ignored, the equilibrium order quantity is the same as the traditional newsvendor case, wholesale price is smaller, and the supplier earns lower profit.*

Proposition 10.6 only obtains the equilibrium solutions $(\hat{w}_b^*, \hat{q}_b^*)$ in the Bankruptcy Region. Note that we also have the Traditional Newsvendor Region, in which $(\bar{w}_s^*, \bar{q}_s^*)$ can be optimal, and the Ordering with All Wealth Region, in which (w^2, q^2) can be optimal (note that in the Ordering with All Wealth Region, the supplier wants the retailer to have the smallest order quantity). Therefore, we have three candidates of the global equilibrium solutions: $(\hat{w}_b^*, \hat{q}_b^*)$, $(\bar{w}_s^*, \bar{q}_s^*)$, and (w^2, q^2). According to the relationships among $\hat{q}_b^*$, $\bar{q}_s^*$, and q^2, we have the following proposition.

Proposition 10.7 *The global equilibrium solutions, (w_b^{**}, q_b^{**}), are:*

1. $(\bar{w}_s^*, \bar{q}_s^*)$ *if* $\bar{q}_s^* \leq q^2$
2. (w^2, q^2) *if* $\bar{q}_s^* > q^2$ *and* $\hat{q}_b^* < q^4$
3. *either* (w^2, q^2) *or* $(\hat{w}_b^*, \hat{q}_b^*)$, *whichever gives the supplier a larger profit, if* $\bar{q}_s^* > q^2$ *and* $\hat{q}_b^* \geq q^4$

From Proposition 10.7, we have $q_b^{**} \leq \bar{q}_s^*$. Then, the efficiency of decentralized supply chain might decrease due to the existence of the bankruptcy costs.

We now provide some sensitivity analysis results. The analysis is tractable only for the cases where the production cost c is small and can be ignored. Then, (10.12) is reduced to $1 - q_b z(q_b) = 0$, which implies that $\tilde{q}$ is always the optimal order quantity to the supplier.

Lemma 10.5. *In the Bankruptcy Region, when $c = 0$, the supplier's equilibrium wholesale price w_b^{**} and profit $\Pi_b^*(q_b^{**})$ decrease in α_1, α_2, and B. Also, the supplier's equilibrium wholesale price and profit increase in y. Finally, the supplier's equilibrium wholesale price and profit increase in x if $\alpha_1 \geq \alpha_2$.*

Lemma 10.5 satisfies our intuition since the supplier's equilibrium wholesale price and profit decrease in the bankruptcy costs. Also, since the equilibrium order quantity is fixed at $\tilde{q}$, the more working capital the retailer has (the collateral value x is fixed), the smaller the Bankruptcy Region is. Then, the supplier can charge a larger wholesale price and get higher profits. Similarly, we can explain why the supplier's equilibrium wholesale price and profit increase in x (y is fixed). From our discussion, the supplier has a preference for working with rich retailers in the supply chain.

We would like to emphasize that both the equilibrium order quantity and wholesale price increase in the retailer's wealth. When the retailer is poor, the bank charges a high interest rate due to the bankruptcy costs, which are eventually reflected in the purchasing cost in addition to the wholesale price. In this situation, the retailer's equilibrium order quantity increases in his wealth, since the bankruptcy costs decrease in his wealth. On the other hand, the supplier can also share the benefit resulting from the decrease of the retailer's bankruptcy costs and end up with a larger wholesale price.

10.4 Financing the Bankrupt-Prone Newsvendor with Trade Credit Contracts

In the case of being financially constrained, small size and startup firms often have difficulty accessing financing from banks, due to their lack of collateral, lack of credit history, and the tenuous nature of their business establishment (Vandenberg, 2003). Supplier financing of retail inventory, often referred to as "trade credit," has become the prevailing short-term financing source for buyers even for environments with very well-developed financial markets, such as the U.S. (Petersen and Rajan 1997). In less-developed countries where formal financial institutions are scarce and less effective, trade credit plays an even more significant role (Fisman and Love 2003).

The trade credit literature is broad and multidisciplinary, and for broad reviews on economic rationale and financial theories of it see Fisman and Love (2003), Burkart and Ellingsen (2004), Giannetti et al. (2007), and references therein. Theoretical models in the economics and finance literature typically emphasize information asymmetry among the players and other features of trade credits, such as price discrimination, differentiated goods, long-term customer relationships, and low diversion value of supplier products, to explain why trade credits are prevailing. However, they fail to model operational considerations within a supply chain setting (e.g. ordering decisions in response to uncertain demand). Our work will provide a supply chain theory based rationale for the effectiveness of trade credit contracts by explicitly modeling the operational interactions of the supplier and retailer, and thus offers an alternative explanation and a perspective missing in the current trade credit literature.

Short-term financing within a "selling to the bankrupt-prone newsvendor" setting appears in recent working papers of Gupta (2008) and Zhou and Groenevelt (2007). Gupta (2008) compares supplier financing with bank financing. In his paper, the wholesale price is exogenous and only the early payment discount rate is the supplier's decision variable. Zhou and Groenevelt (2007) analyze two asset based financing schemes with credit line limits: a supplier subsidized bank financing (e.g., the retailer receives a bank loan where the interest is paid by the supplier) and open account financing (e.g., trade credit with no early payment discount). Since the retailer only pays the loan principal in both schemes, the supplier's trade credit contract has only one parameter: the wholesale price. Our work jointly optimizes the supplier's wholesale price and interest rate, and the obtained results provide new insights on the supplier's rationale behind trade credit practices and the prevalence of open account financing. For more detail of Section 10.4, please refer to Kouvelis and Zhao (2008).

10.4.1 DESCRIPTION OF EARLY PAYMENT DISCOUNT TRADE CREDIT MODEL

10.4.1.1 Sequence of Events. We consider a supply chain where a retailer has a single opportunity to order "now" a product from a supplier to satisfy future uncertain demand. The supplier offers a take-it-or-leave-it trade credit contract (w, r_s) to the retailer. If the retailer accepts the contract, he places the order, and can either pay now at the wholesale price w, or pay at the higher price $w(1 + r_s)$ at the end of the sales season. Then, r_s is the supplier's interest rate for the period from now to the end of sales season. Such a trade credit contract is referred to as "supplier early payment discount" or "supplier financing" for simplicity. The case $r_s = 0$ is referred to as "open account financing."

If r_s is large, the retailer might prefer to pay the whole order now by using bank financing for his ordering needs. The interest rate of the bank loan for the period from now to the end of the sales season is denoted by r_b (note: it is a function of the retailer's order quantity and supplier's wholesale price, see Sections 10.3.2.1 and 10.3.2.2). Alternatively, if r_s is small, the retailer prefers to delay the payment to the end of sales season by choosing supplier financing.

The sequence of events is shown in Table 10.3. From the table, we can see there are two time points considered: "now" and "the end of sales season." All decisions are made now, in the order listed in the table: The supplier first decides on the offered contract (w, r_s) by predicting the retailer's responses, and then the retailer decides on the order quantity to place and how to finance it. The bank decides the offered interest rate to the supplier and/or retailer when approached for a loan.

TABLE 10.3 Sequence of Events and Decision Protocol

	Supplier Announces the Trade Credit Contract (w, r_s)	
	Bank Financing (r_s is large)	**Supplier Financing (r_s is small)**
Now	Retailer decides his order quantity	Retailer decides his order quantity
	Retailer borrows a bank loan	
	Bank decides the interest rate	Retailer pays the supplier *in part*
	Retailer pays the supplier *in full*	Supplier borrows a necessary bank loan
		Bank decides the interest rate
End of the sales season	Retailer repays bank loan obligation	Retailer repays trade credit loan obligation
		Supplier repays bank loan obligation

10.4.1.2 Notation and Assumptions. We use subscript s for supplier financing and b for bank financing. Also, we use superscript $*$ to denote variables when the retailer selects the optimal order quantity for given contract parameters, and superscript $**$ to denote the supplier's optimal contract parameters in the Stackelberg equilibrium. For example, let q be the retailer's order quantity. Then, q_s is the retailer's order quantity under supplier financing, $q_s^*(w, r_s)$ is his optimal order quantity for given w and r_s, and q_s^{**} is his Stackelberg equilibrium order quantity for the equilibrium contract parameters w^{**} and r_s^{**}.

Our other modeling assumptions are A2 through A6 listed in Section 10.2, and assumption A1 is replaced with the following A1′.

A1′ : The bank, retailer and supplier are all risk neutral, and the market is perfect (no taxes, transaction costs, and no bankruptcy costs) with full information.

From assumption A1′, $\alpha_1 = \alpha_2 = B = 0$, i.e., no bankruptcy costs is considered, in the whole Section 10.4. Also, we remark that it is an implicit assumption behind our work that a supplier is willing to extend credits to her retailers even if she has to borrow banks loan to cover the production requirements. This assumption is justified by our discussion in Section 10.4.3.2, which clearly implies that in the presence of competitively priced bank loans, the supplier's capital constraint does not affect her trade credit contract decisions. The realism of the practice is confirmed in Horen (2007).

10.4.2 RETAILER'S PERSPECTIVE: SUPPLIER OR BANK FINANCING

10.4.2.1 Retailer's Problem under Bank Financing (Benchmark). The retailer chooses bank financing in order to take advantage of the low wholesale price w (lower than the price $w(1 + r_s)$ for the delayed payment). Then, the trade credit contract parameter r_s plays no role in bank financing.

Since $\alpha_1 = \alpha_2 = B = 0$, we have $G(\theta) \equiv 1$ for any $\theta \geq 0$ where $G(\theta)$ is defined in (10.8). From $B = 0$, we have $b \leq 0$. Then, (10.7) can be simplified as

$$\Gamma(k_b(w, q_b)) = p\mathbb{E}(\min\{\xi, k_b(w, q_b)\}) + x. \tag{10.16}$$

Consequently, from (10.4), we have that

$$p\mathbb{E}(\min\{\xi, k_b(w, q_b)\}) = wq_b(1 + r_f) - x - y(1 + r_f) = wq_b(1 + r_f) - \Omega.$$

Therefore, for $q_b \geq q_x(w)$, from (10.2a):

$$\begin{aligned}\pi_b(w, q_b) &= p\mathbb{E}[\min\{\xi, q_b\}] - p\mathbb{E}(\min\{\xi, k_b(w, q_b)\}) && (10.17)\\ &= p\mathbb{E}[\min\{\xi, q_b\}] - wq_b(1 + r_f) + \Omega && (10.18)\end{aligned}$$

This is identical to (10.2b) for $0 \leq q_b \leq q_x(w)$. We refer to the model that optimizes (10.18) with respect to q_b as the bank financing model. Following standard literature terminology, we also refer to it as the *unconstrained newsvendor* model with the time value of the investment explicitly considered (i.e., the so-called "traditional newsvendor" in Section 10.3).

Furthermore, from Proposition 10.1 in Section 10.3.2.3, we have $\Omega_1(w) = \Omega_2(w) = \Omega_3(w) = w(1+r_f)\bar{F}^{-1}\left(\frac{w(1+r_f)}{p}\right)$, and for any $\Omega \geq 0$, we have:

$$q_b^*(w) = \bar{F}^{-1}\left(w(1+r_f)/p\right) \tag{10.19}$$

This defines the retailer's response curve under bank financing in (w, q) space. This response curve is also the response curve of the unconstrained newsvendor. Note that in Section 10.3.3.1, we show that due to the bankruptcy costs, the retailer's optimal order quantity is less than or equal to the unconstrained newsvendor model (please refer to Figure 10.3). Now, since there are no bankruptcy costs, the retailer's optimal order quantity is always equal to the unconstrained newsvendor order quantity.

Let $r_b^*(w) = r_b(w, q_b^*(w))$, $k_b^*(w) = k_b(w, q_b^*(w))$ and $\pi_b^*(w) = \pi_b(w, q_b^*(w))$ be the bank's interest rate, the retailer's bankruptcy threshold, and his expected terminal cash flow, respectively, when the retailer orders $q_b^*(w)$. Then, for $k_b^*(w) > 0$, (10.17) can be rewritten as:

$$\pi_b^*(w) = p\mathbb{E}[\min\{\xi, q_b^*(w)\}] - p\mathbb{E}[\min\{\xi, k_b^*(w)\}] \tag{10.20}$$

Note that (10.19) implies that the optimal order quantity is not influenced by the financial constraint. The result is consistent with the Modigliani and Miller (1958) (M&M theory) results, and with results in Xu and Birge (2004). We summarize the result in Proposition 10.8.

Proposition 10.8 *The capital constrained retailer under bank financing with competitively priced loan and under perfect market assumptions separates ordering and financing decisions. He orders $q_b^*(w)$ like an unconstrained newsvendor as in (10.19) and then borrows the needed amount $wq_b^*(w) - y$.*

10.4.2.2 Retailer's Problem under Supplier Financing. We now study the retailer's problem under supplier financing. First, we analyze the retailer's terminal cash flows and define his objective functions. Then, we find his optimal quantity $q_s^*(w, r_s)$, and establish the 1-to-1 relationship between $q_s^*(w, r_s)$ and w for a given r_s. We focus on $r_s \geq r_f$ throughout this paper (we explicitly address the case $r_s < r_f$ in Section 10.4.3.4).

We first analyze the retailer's terminal cash flows. There are two cases: the retailer's working capital is enough to pay the order up front at the wholesale price w, i.e., $y \geq wq_s$, and otherwise, $y < wq_s$.

When $y \geq wq_s$, the retailer pays the supplier wq_s up front, and invests the rest, $y - wq_s$, at the risk-free rate r_f. At the end of sales season, he receives sales revenue, $p\min\{\xi, q_s\}$, the collateral assets, x, and his capital investment returns, $(y - wq_s)(1+r_f)$. The retailer's net cash flow is:

$$p\min\{\xi, q_s\} - wq_s(1+r_f) + x + y(1+r_f) \tag{10.21}$$

When $y < wq_s$, the retailer pays the supplier y up front and delays the rest, $wq_s - y$, at the interest rate r_s. At the end of sales season, he receives sales revenue, $p \min\{\xi, q_s\}$, and collateral assets, x, and repays his loan obligation, $(wq_s - y)(1 + r_s)$, to the extent possible. Then, his net cash flow is:

$$[p \min\{\xi, q_s\} + x - (wq_s - y)(1 + r_s)]^+ \tag{10.22}$$

If $wq_s - y(1 + r_s) \leq x$, the retailer can definitely repay the loan and get non-negative net cash flow. Otherwise, let $k_s(w, q_s, r_s)$ be the retailer's *bankruptcy threshold* under supplier financing that is:

$$k_s(w, q_s, r_s) = \frac{1}{p}[(wq_s - y)(1 + r_s) - x]^+$$

If $w(1 + r_s) \geq p$, trade credits are not profitable for the retailer. Then, the retailer accepts trade credits only if $w(1 + r_s) < p$, which implies that $k_s(w, q_s, r_s) < q_s$ whenever there exists trade credit transaction between the supplier and retailer.

We summarize the retailer's expected cash flows at the end of sales season as follows:

$$\pi_s(w, q_s, r_s) = \begin{cases} p\mathbb{E}[\min\{\xi, q_s\}] - p\mathbb{E}[\min\{\xi, k_s(w, q_s, r_s)\}], & \text{if } wq_s > x/(1 + r_s) + y \quad (10.23a) \\ p\mathbb{E}[\min\{\xi, q_s\}] - wq_s(1 + r_s) + x + y(1 + r_s), & \text{if } y < wq_s \leq x/(1 + r_s) + y \quad (10.23b) \\ p\mathbb{E}[\min\{\xi, q_s\}] + x, & \text{if } wq_s = y \quad (10.23c) \\ p\mathbb{E}[\min\{\xi, q_s\}] - wq_s(1 + r_f) + \Omega, & \text{if } wq_s < y \quad (10.23d) \end{cases}$$

The four cases corresponding to (10.23d), (10.23c), (10.23b), and (10.23a) are referred to as the retailer's *Partial Working Capital Use* region, *All Working Capital Use Region*, *Borrowing/No Bankruptcy Region*, and *Bankruptcy region*, which we can establish to be $[0, q_3^l] \cup [q_3^u, +\infty)$, $[q_3^l, q_2^l(r_s)] \cup [q_2^u(r_s), q_3^u]$, $[q_2^l(r_s), q_1^l(r_s)] \cup [q_1^u(r_s), q_2^u(r_s)]$, and $[q_1^l(r_s), q_1^u(r_s)]$, respectively. The wholesale prices on the response curve corresponding to $q_i^j(r_s)$ and q_3^j are $w_i^j(r_s) = \frac{p\bar{F}(q_i^j(r_s))}{1+r_s}$ and $w_3^j = \frac{p\bar{F}(q_3^j)}{1+r_f}$ for $j = l, u$ and $i = 1, 2$, respectively. Please refer to Figure 10.5.

Given w and r_s, the retailer's optimal order quantity $q_s^*(w, r_s)$ is or can be solved from

$$\begin{cases} p\bar{F}(q_s^*(w, r_s)) = w(1 + r_s) \bar{F}(k_s(w, q_s^*(w, r_s), r_s)) & \text{if } q_s^*(w, r_s) \in [q_1^l(r_s), q_1^u(r_s)] \quad (10.24a) \\ q_s^*(w, r_s) = \bar{F}^{-1}(u(1+r_s)/p) & \text{if } q_s^*(w, r_s) \in U_1(r_s) \quad (10.24b) \\ q_s^*(w, r_s) = y/w & \text{if } q_s^*(w, r_s) \in U_2(r_s) \quad (10.24c) \\ q_s^*(w, r_s) = \bar{F}^{-1}(u(1+r_f)/p) & \text{if } q_s^*(w, r_s) \in [0, q_3^l] \cup [q_3^u, +\infty) \quad (10.24d) \end{cases}$$

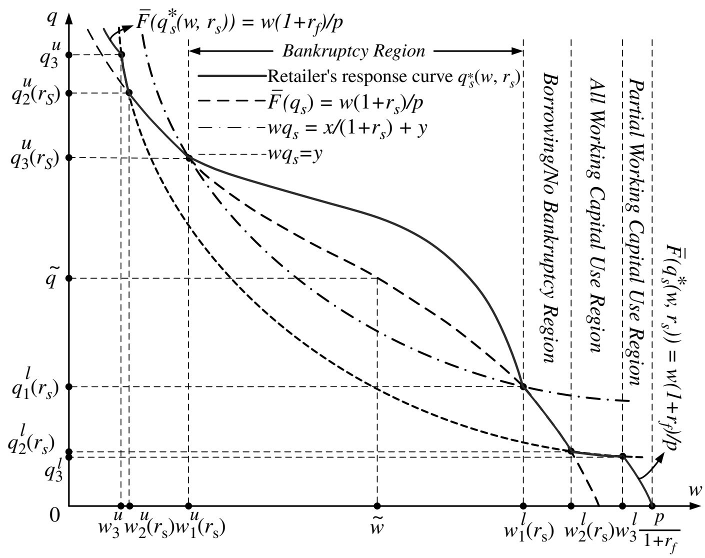


FIGURE 10.5 The response curve $q_s^*(w, r_s)$ for a given r_s.

Note that $U_1(r_s) = [q_2^l(r_s), q_1^l(r_s)] \cup [q_1^u(r_s), q_2^u(r_s)]$ and $U_2(r_s) = [q_3^l, q_2^l(r_s)] \cup [q_2^u(r_s), q_3^u]$. We refer to $q_s^*(w, r_s)$ as the retailer's response curve under supplier financing.

In the Bankruptcy Region, $q_s^*(w, r_s)$ obtained from (10.24a) is a function of w and r_s. That is, the optimal order quantity depends on the financial term r_s, which implies the need for integrated financing and inventory decisions. Since r_s is a single value decision variable and not a function of the risk associated with the loan, the supplier's trade credit loan does not competitively price demand risks. In a decentralized supply chain operating under a trade credit contract, with the retailer and supplier acting as independent agents, the emerging "agency issues" (see Barnea et al. 1981 for relevant discussion) violate the M&M theory premises and lead to the interdependency of optimal financing and inventory decisions.

For a given r_s, the retailer's response curve $q_s^*(w, r_s)$ in (w, q) space is shown as the solid line in Figure 10.5. For $q_s^*(w, r_s) \in [0, q_3^l] \cup [q_3^u, \infty)$, either w or $q_s^*(w, r_s)$ is small so that the retailer's working capital is enough for the order. Then, his marginal cost of each additional item is $w(1 + r_f)$, and his response follows the unconstrained newsvendor curve $q_s^*(w, r_s) = \bar{F}^{-1}(w(1 + r_f)/p)$, which is also his response curve under bank financing (please refer to (10.19)). For $q_s^*(w, r_s) \in [q_3^l, q_2^l(r_s)] \cup [q_2^u(r_s), q_3^u]$, the retailer's response curve is $q_s^*(w, r_s) = y/w$ since he orders with all working capital. For $q_s^*(w, r_s) \in [q_2^l(r_s), q_1^l(r_s)] \cup [q_1^u(r_s), q_2^u(r_s)]$, the retailer has to borrow, but does not face bankruptcy risks, due to the collateral assets x. Then, his marginal cost of each additional item is $w(1 + r_s)$, and his response follows the unconstrained newsvendor curve $q_s^*(w, r_s) = \bar{F}^{-1}(w(1 + r_s)/p)$.

Finally, for $q_s^*(w, r_s) \in [q_1^l(r_s), q_1^u(r_s)]$, the loan the retailer borrows cannot be secured by x, and he faces bankruptcy risks. If the retailer bankrupts, he loses his collateral and realized sales, which are smaller than his loan obligations. That is, his expected marginal cost is less than $w(1 + r_s)$. As a result, the retailer places a larger order quantity to take advantage of his limited liability, and his response curve ends up being above $q_s^*(w, r_s) = \bar{F}^{-1}(w(1 + r_s)/p)$.

Let $k_s^*(w, r_s) = k_s(w, q_s^*(w, r_s), r_s)$ and $\pi_s^*(w, r_s) = \pi_s(w, q_s^*(w, r_s), r_s)$ be the retailer's bankruptcy threshold and expected terminal cash flow, respectively, when the retailer orders $q_s^*(w, r_s)$.

10.4.2.3 Choice of Financing: Supplier or Bank?. Note that the common parameter under bank financing and supplier financing is the wholesale price w. For each w, we first consider the case that the supplier takes $r_b^*(w)$ as her interest rate r_s (i.e., the supplier's interest rate is fixed at the bank's interest rate).

In this case, our detailed analysis indicates that even if the supplier's interest rate equals the bank's, the retailer still selects supplier financing, and orders a larger quantity and receives larger expected cash flow. An intuitive explanation is that under bank financing, the retailer's optimization problem is restricted by the competitively priced loan equation (10.4). However, under supplier financing, the supplier's preset interest rate is not a function of the retailer's order quantity. Thus, the retailer's optimization problem is unrestricted, with the resulting optimum order quantity and expected cash flows larger than those under bank financing.

As shown in Figure 10.6, the thick and thin solid lines in (w, q_s) space are the retailer's response curves under bank and supplier financing, respectively. For

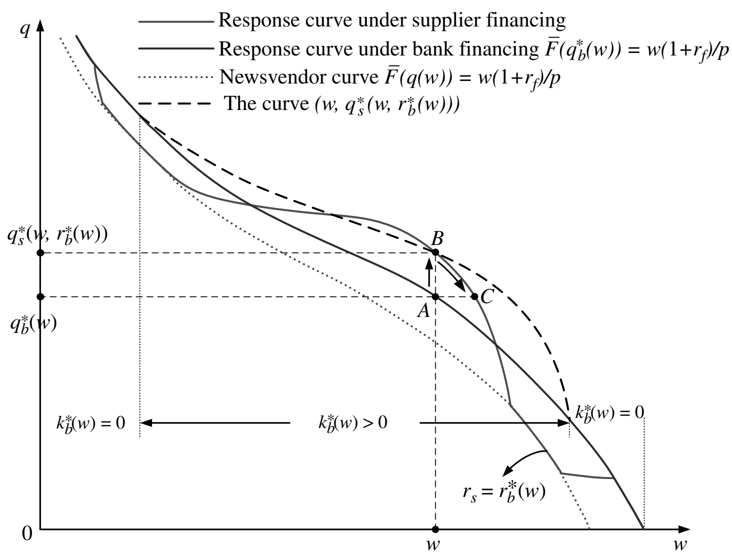


FIGURE 10.6 Retailer orders more under supplier financing.

a given w, A and B are two response points under bank and supplier financing, respectively, and the order quantity at B is larger than at A. At A, both bank and supplier financing yield the same expected cash flows, based on (10.20), (10.23a), and the fact that $k_b(w, q_b^*(w)) = k_s(w, q_b^*(w), r_b^*(w))$. Then, under supplier financing, the retailer's expected cash flow at B is better than at A since B is the optimal response point.

Next, Lemma 10.6 (see below) establishes that for any given w, the retailer will choose supplier financing for any $r_f \leq r_s \leq r_b^*(w)$, by showing that for a given w, the retailer's optimal expected cash flow under supplier financing decreases in r_s. In Section 10.4.3.3, we establish that $r_s^{**} = r_f$ is the optimal trade credit rate set by the supplier. Then, Lemma 10.6 proves that when the supplier chooses her optimal interest rate, the retailer always prefers supplier to bank financing.

Lemma 10.6. *For a given w, and for any $r_f \leq r_s \leq r_b^*(w)$, the retailer's expected cash flow under supplier financing is no less than that under bank financing, and thus he will choose supplier financing.*

10.4.3 SUPPLIER'S PERSPECTIVE: OPTIMAL CONTRACT PARAMETERS

10.4.3.1 Supplier's Problem under Bank Financing (Benchmark). The supplier's problem under bank financing is discussed in Section 10.3.3.2. Recall that the supplier's expected terminal cash flow is:

$$\Pi_b^*(q_b) = (w_b^*(q_b) - c)q_b(1 + r_f) \tag{10.25}$$

The discussion in Section 10.4.2.1 implies that the retailer's expected terminal cash flow models are identical with or without bankruptcy risks. As a result, the discussion for the case without bankruptcy costs (i.e., $G(\theta) \equiv 1$ for any $\theta \geq 1$) in Section 10.3.3.2 can be applied to the case that the retailer does not have bankruptcy risks. In other words, $(\hat{w}_b^*, \hat{q}_b^*)$ are the *global* Stackelberg equilibrium solutions (i.e., $w_b^{**} = \hat{w}_b^*$ and $q_b^{**} = \hat{q}_b^*$). Recall that $\hat{q}_b^*$ can be solved from (10.15) and then $\hat{w}_b^*$ can be solved from (10.13). Let $r_b^{**} = r_b^*(w_b^{**})$ be the bank's equilibrium interest rate.

10.4.3.2 Supplier's Problem under Supplier Financing. For a fixed $r_s \geq r_f$, we can show that there is a 1-to-1 mapping between w and the retailer's optimal order quantity $q_s^*(w, r_s)$. Let $w_s^*(q_s, r_s)$ be the inverse function of $q_s^*(w, r_s)$. That is, for any $q_s \geq 0$, if the supplier's wholesale price is $w_s^*(q_s, r_s)$, then q_s will be the retailer's optimal order quantity. Due to the 1-to-1 relationship, we will use (q_s, r_s), instead of (w, r_s), as the basic contract parameters (see Lariviere and Porteus, 2001, for similar arguments). In Section 10.4.2.2, we denote $k_s^*(w, r_s) = k_s(w, q_s^*(w, r_s), r_s)$ and $\pi_s^*(w, r_s) = \pi_s(w, q_s^*(w, r_s), r_s)$. Now, we use the relevant for our further discussion notation $k_s^*(q_s, r_s) = k_s(w_s^*(q_s, r_s), q_s, r_s)$ and $\pi_s^*(q_s, r_s) = \pi_s(w_s^*(q_s, r_s), q_s, r_s)$.

Let $M = \min\{y, w_s^*(q_s, r_s)q_s\}$ be the amount the supplier receives from the retailer now. Also, let the amount she receives from the retailer at the end of sales season be $N(\xi) = \min\{p\min\{\xi, q_s\} + x, (w_s^*(q_s, r_s)q_s - M)(1 + r_s)\}$, where the first term in the outside minimum is the retailer's wealth if he bankrupts and the second term is the loan principal and interest if the retailer has enough wealth. Similar as our analysis of the retailer's problem under bank financing in Section 10.4.2.1, we can show that under competitively priced bank loans and without bankruptcy costs, the supplier's working capital Y and collateral assets X do not affect her decisions on the trade credit contract. That is, we can ignore X and Y from the supplier's problem without affecting the optimal solutions (we have done so in Section 10.4.3.1). Thereafter, we assume $X = Y = 0$, and the supplier's expected terminal cash flow is:

$$\Pi_s^*(q_s, r_s) = \mathbb{E}[N(\xi)] + (M - cq_s)(1 + r_f) \tag{10.26}$$

If $w_s^*(q_s, r_s)q_s \le y$ (i.e., $q_s \in [0, q_2^l(r_s)] \cup [q_2^u(r_s), +\infty)$, then the retailer has paid the order upfront so that $M = w_s^*(q_s, r_s)q_s$ and $N(\xi) = 0$. Alternatively, if $y < w_s^*(q_s, r_s)q_s$ (i.e., $q_s \in [q_2^l(r_s), q_2^u(r_s)]$), then $M = y$ and $N(\xi) = \min\{p\min\{\xi, q_s\} + x, (w_s^*(q_s, r_s)q_s - y)(1 + r_s)\}$. Note that $k_s^*(w_s^*(q_s, r_s), q_s, r_s) = \frac{1}{p}[(w_s^*(q_s, r_s)q_s - y)(1 + r_s) - x]^+$. After simplification, we obtain the supplier's expected cash flow as:

$$\Pi_s^*(q_s, r_s) = \begin{cases} p\mathbb{E}[\min\{\xi, k_s^*(q_s, r_s)\}] - cq_s(1 + r_f) + \Omega, \\ \qquad \text{if } q_s \in [q_1^l(r_s), q_1^u(r_s)] & (10.27\text{a}) \\ w_s^*(q_s, r_s)q_s(1 + r_s) - cq_s(1 + r_f) - y(r_s - r_f), \\ \qquad \text{if } q_s \in [q_2^l(r_s), q_1^l(r_s)] \cup [q_1^u(r_s), q_2^u(r_s)] & (10.27\text{b}) \\ w_s^*(q_s, r_s)q_s(1 + r_f) - cq_s(1 + r_f) \\ \qquad \text{if } q_s \in [q_3^l, q_2^l(r_s)] \cup [q_2^u(r_s), q_3^u] & (10.27\text{c}) \\ w_s^*(q_s, r_s)q_s(1 + r_f) - cq_s(1 + r_f) \\ \qquad \text{if } q_s \in [0, q_3^l] \cup [q_3^u, +\infty) & (10.27\text{d}) \end{cases}$$

10.4.3.3 The Optimal Trade Credit Contract Parameters. Our analysis in Sections 10.4.3.3 to 10.4.4.2 assumes demand distributions with increasing (IFR) and convex failure rates. Then, we have the following Proposition 10.9.

Proposition 10.9 *Let $r_s \ge r_f$. For demand distributions with increasing and convex failure rates, the supplier's Stackelberg equilibrium interest rate for the trade credit contract is $r_s^{**} = r_f$.*

We intuitively explain why $r_s^{**} = r_f$ is optimal to the supplier. If w is close to the production cost c, then up-front paid units are not profitable for the supplier, and if w is close to $\frac{p}{1+r_f}$, then the retailer's small order quantity is far from optimal for the supplier. As a result, the optimal w is close to the middle of $[c, \frac{p}{1+r_f}]$. Let $W = w(1 + r_s)$, which is the wholesale price for delayed payment. Similarly, the optimal $\frac{W}{1+r_f}$ is also close to the middle of $[c, \frac{p}{1+r_f}]$. If r_s is large, w and $\frac{W}{1+r_f}$ cannot be both close to the middle of $[c, \frac{p}{1+r_f}]$. An extreme example is $r_s = \frac{p}{c} - 1$. In this case, the only feasible solution to the supplier is $w = c$ (and $W = w(1 + r_s) = p$), which leaves her with zero profit. Thus, she tends to decrease r_s, and theoretically, we show the optimal value is r_f.

Proposition 10.9 implies that as r_s decreases, the retailer's bankruptcy threshold becomes larger (i.e., the risk neutral retailer orders larger quantities and is willing to bear more risk due to his limited liability). On the other hand, the risk neutral supplier continues to finance such risk as she acts as a bond holder on the retailer's cash flow. In such a system it is efficient to encourage the retailer to take risks by lowering the borrowing rates, while the supplier compensates for the cheap trade credit by charging a larger wholesale price than an equivalent wholesale price contract that would have incentivized the same retailer's order quantity.

At $r_s^{**} = r_f$, (10.24b) and (10.24d) are identical. Then, under supplier financing, the three regions other than the Bankruptcy Region merge into one, which we refer to as the *Non-Bankruptcy* Region. The boundary point of Bankruptcy and Non-Bankruptcy Regions is $(w_1^l(r_s^{**}), q_1^l(r_s^{**}))$. Note that in the Non-Bankruptcy Region, the retailer's response curves under supplier and bank financing are identical. Please refer to Figure 10.7.

We next study the equilibrium order quantity q_s^{**}, as a function of the retailer's wealth $\Omega = x + y(1 + r_f)$. Let $\hat{q}_s^*(r_s)$ be the solution to:

$$\frac{q_s f(q_s) - \bar{F}(q_s)}{1 - \frac{w_s^*(q_s, r_s)(1+r_s)}{p} q_s z(k_s^*(q_s, r_s))} + \frac{c(1 + r_f)}{p} = 0 \tag{10.28}$$

Recall that $\bar{q}_s^*$ is the solution of (10.15) (please refer to discussion in Section 10.3.3.2), and it is not a function of r_s. Also, $\bar{q}_s^* = q_b^{**}$, where q_b^{**} is the equilibrium order quantity under bank financing (please refer to the discussion in Section 10.4.3.1).

Our detailed analysis establishes that under supplier financing, the equilibrium order quantity q_s^{**} is either $\hat{q}_s^*(r_s^{**})$ or $\bar{q}_s^*$, depending on the retailer's wealth level Ω. For demand distributions with failure rates both increasing and convex, we show that there exist $0 < \underline{\Omega} \leq \bar{\Omega}$. The case that $\Omega < \underline{\Omega}$ is referred to as the retailer is "poor." In this case, $\bar{q}_s^* > q_1^l(r_s^{**})$ cannot be optimal, and the supplier's optimal solution is $\hat{q}_s^*(r_s^{**})$, which is in the retailer's Bankruptcy Region. Next, $\bar{\Omega} \geq \Omega \geq \underline{\Omega}$ is interpreted as the retailer is "moderately rich." In this case, $\hat{q}_s^*(r_s^{**}) \geq q_1^l(r_s^{**}) \geq \bar{q}_s^*$, and the optimal solution can be either in the retailer's Bankruptcy or in Non-Bankruptcy Region, whichever gives the supplier the larger expected cash flow. Finally, $\Omega > \bar{\Omega}$ is interpreted as the retailer is "rich" and does not need financing. In this case, $\hat{q}_s^*(r_s^{**}) < q_1^l(r_s^{**})$ cannot be optimal. Then, the optimal solution is $\bar{q}_s^*$, which is in the retailer's Non-Bankruptcy Region, identical to the case that the retailer has no financial constraints. The wholesale price $w_s^{**} = w_s^*(q_s^{**}, r_s^{**})$ can be determined from q_s^{**} correspondingly. Please refer to the three cases in Figure 10.7.

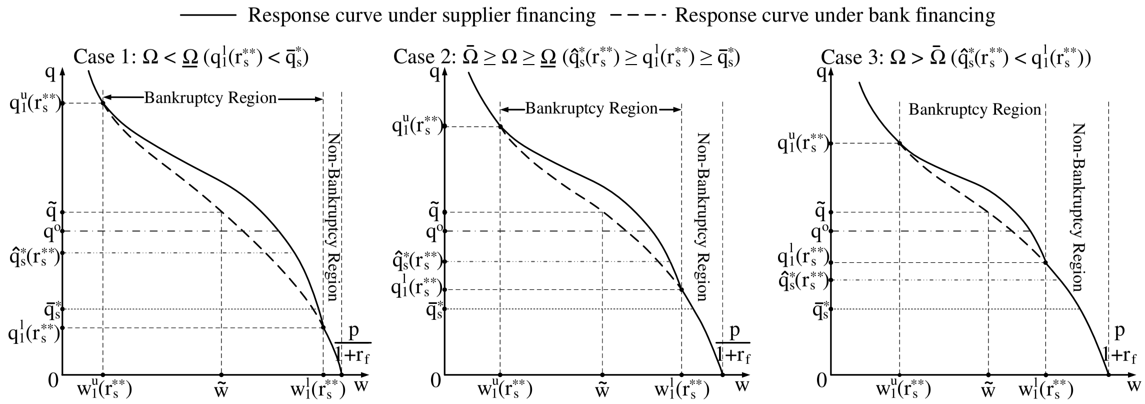


FIGURE 10.7 **Retailer's three wealth levels for equilibrium order quantities.**

Under bank financing, the bank does not share demand risks with the retailer, due to the nature of the competitively priced loans. As a result, the retailer faces all demand risks. However, under supplier financing, the supplier partially shares demand risks with the retailer through the lower trade credit rates. This risk-sharing mechanism improves the supply chain efficiency but not at the level to fully coordinate the supply chain. The result is summarized in the following proposition.

Proposition 10.10 *Supplier financing improves the supply chain efficiency over bank financing, but cannot coordinate the chain.*

10.4.3.4 The Supplier's Interest Rate is Less Than the Risk-Free Rate. In this section, we briefly discuss the case of $r_s \leq r_f$ under supplier financing. Then, the retailer optimally decides to invest his working capital y at the risk-free rate r_f, and delays the payment to the supplier. At the end of sales season, he receives sales revenue, $p \min\{\xi, q_s\}$, collateral assets, x, returns on his capital investment, $y(1 + r_f)$, and repays the supplier's loan, $wq_s(1 + r_s)$, to the extent possible. Then, the retailer's and supplier's expected terminal cash flows are

$$\pi_s(w, q_s, r_s) = \mathbb{E}[p \min\{\xi, q_s\} + \Omega - wq_s(1 + r_s)]^+ \quad (10.29)$$

$$\Pi_s(w, q_s, r_s) = \mathbb{E}[\min\{p \min\{\xi, q_s\} + \Omega, wq_s(1 + r_s)\}] - cq_s(1 + r_f) \quad (10.30)$$

We have established for $r_s \geq r_f$, $r_s^{**} = r_f$. For $0 \leq r_s \leq r_f$, we have the following proposition.

Proposition 10.11 *For $0 \leq r_s \leq r_f$, any (w, r_s) satisfying $w(1 + r_s) = w_s^{**}(1 + r_s^{**})$ is an optimal pair of contract parameters and gives the same expected cash flows to the supplier and retailer as (w_s^{**}, r_s^{**}).*

Let $W^{**} = w_s^{**}(1 + r_s^{**})$ be the optimal wholesale price for delayed payment where $r_s^{**} = r_f$. If we fix the delayed wholesale price at W^{**}, then for any r_s in the range $[0, r_s^{**})$, the wholesale price w (i.e., $w = \frac{W^{**}}{1+r_s}$) fails to motivate the retailer to pay early. Then, both his optimal order quantity and expected terminal cash flow are not affected by r_s, and the supplier's expected terminal cash flow is also not affected by r_s. As a result, the results of contract (w, r_s) are identical to those of (w_s^{**}, r_s^{**}). This explains why any interest rates in the region $0 \leq r_s \leq r_f$ are also optimal.

10.4.4 SUPPLIER FINANCING AND THE SUPPLY CHAIN

10.4.4.1 Supplier's Improved Profitability via Financing the Retailer. In this section, we compare the supplier's optimal expected cash flow under supplier financing with that under bank financing, and answer the question whether she should finance the retailer or not.

Proposition 10.12 *For demand distributions with increasing and convex failure rates, the supplier's optimal expected cash flow under supplier financing is at least as good as that under bank financing.*

We interpret Proposition 10.12 intuitively now. Since in supplier financing, the loan's interest rate is a single value instead of competitively priced. Then, due to the limited liability, the retailer orders a quantity larger than the traditional newsvendor case. The supplier benefits from this retailer's larger order quantity, and gets a better profit.

Proposition 10.12 assumes no credit limits on trade credits. If the supplier does place a limit, the retailer has to partially choose bank loans for his procurement needs. Then, there is less sharing of demand risks between the supplier and retailer, and the supplier's expected cash flow and supply chain efficiency are reduced (but still larger than those under pure bank financing). Thus, a risk neutral wealth optimizing supplier has no incentive to place credit limits on her loans to the retailer.

The observed in practice trade credit limits are mostly motivated by issues of trust in the retailer's commitment and ability to repay his loan obligations. In our model, we assume the retailer is committed to repay his loan obligations to the extent possible (assumption A6) in Section 10.2. If there are information asymmetry considerations on the estimation of market demand between the players, then trade credit limits might play a role. However, it is outside the scope of our current work, but worthwhile to pursue as a future research avenue.

10.4.4.2 Retailer May Improve Profitability under Supplier Financing.

Proposition 10.10 implies that at equilibrium, the supply chain efficiency is larger under supplier financing than under bank financing. From Proposition 10.12, the supplier always benefits from the improvement of the supply chain efficiency. But can the retailer benefit from the efficiency improvement as well? We will answer this question in this section.

We can establish that for a retailer who is moderately rich (i.e., $\underline{\Omega} \leq \Omega \leq \bar{\Omega}$), his expected cash flow under supplier financing is at least as good as that under bank financing, since he can get a smaller equilibrium wholesale price under supplier financing. Similar result holds even if the retailer is poor but not very poor (i.e., $\underline{\underline{\Omega}} \leq \Omega \leq \underline{\Omega}$) because both the equilibrium wholesale price and interest rate under supplier financing are smaller than those under bank financing. However, the "very poor" retailer (i.e., $0 \leq \Omega < \underline{\underline{\Omega}}$) might be worse under supplier financing. More specifically, the retailer's expected cash flows is closely associated with his initial wealth. Especially, when $\Omega \to 0$, under supplier financing, the equilibrium wholesale price tends to $\frac{p}{1+r_s^{**}}$, and as a consequence, the retailer's expected cash flow tends to zero.

Note that our above discussion constitutes in no way contradiction of our result that the retailer always prefers supplier financing over bank financing. Even the "very poor" retailer, when offered an optimally structured trade credit contract, prefers it over bank financing. However, the "very poor" retailer would have preferred the supplier to offer a pure wholesale price contract without any financing provisions, which has a better wholesale price. Then, the retailer would have done his needed financing through the bank. Unfortunately, the supplier as the Stackelberg leader will not offer this contract to him, but an optimally structured trade credit contract.

10.4.4.3 Our Results Are Consistent with Empirical Studies. We predict that trade credit is cheaper than bank credit. Our prediction is consistent with empirical studies across a large sample of firm sizes and industries in different countries. Giannetti et al. (2007) report that in the sample of 3,489 U.S. firms from 1998 National Survey of Small Business Finances, a majority of firms appear to receive trade credits cheaper than bank credit, and even more strikingly, 50% of the most important suppliers offer zero interest rate. After surveying about 1,900 Italian manufacturing firms, Marotta (2005) argues that there is no evidence that trade credit is more expensive than bank credit. Finally, Fabbri and Klapper (2008) document that for over 20% of about 2,500 Chinese firms surveyed, trade credit is cheaper than bank loans. However, we claim that there do exist anecdotes of practices with expensive trade credits, for example "2/10 net 30" contracts have an implied rate of 43.9% (see Ng et al. 1999). The study of motivating factors for such practices is outside the scope of our current work.

10.5 Conclusions and Future Research

In order to effectively model and understand the functioning of financially constrained supply chains, solving the fundamental models of "newsvendor with bankruptcy costs" and "selling to the financially constrained retailer" is essential. Our paper presents solutions for both models under general assumptions about bankruptcy costs. The bankruptcy costs we consider are three types: a portion of sales revenue, depressed collateral value as a portion of the collateral, and fixed administrative costs. The optimal order quantity of the financially constrained retailer is not only a function of the wholesale price, but also depends on the retailer's working capital and collateral. In the two dimensional space of working capital-collateral values, we identified four different zones of ordering behavior of the retailer. Unconstrained retailers order according to our traditional newsvendor formulas. As their wealth decreases, retailers first decrease their order by just ordering what their working capital and collateral allow. Further wealth decreases lead to retailers assuming bankruptcy risks through bank loans. In those cases their order quantities further decrease and the bankruptcy thresholds increase. We would like to note that for the same wealth the order quantities are decreasing in the wholesale price regardless of the ordering behavior and wealth of the retailer.

Our analysis of the supplier's problem of setting wholesale prices when selling to the financially constrained newsvendor establishes the equations for the equilibrium wholesale prices and order quantities in the Bankruptcy Region, and proves that such solutions are unique when the failure rate of the demand distribution is increasing and convex. Our sensitivity analysis indicates that in the Bankruptcy Region, increases in the retailer's wealth lead to increased equilibrium wholesale prices, increased equilibrium order quantities and increased supplier's equilibrium expected profits.

It is clear from our discussion that the presence of costly bankruptcies further decreases the efficiency of the supply chain for a wholesale price contract. However, as the centralized solutions of the supply chain in the presence of bankruptcy costs will depend not only on wholesale prices but also working capital and collateral levels of its constituents, the usual coordinating contracts (e.g., buybacks, revenue sharing, etc.) might not be effective in this setting. In the future, we intend to look into supply chain coordination issues in the presence of bankruptcy costs.

We also present a trade credit model to study the interaction of short-term financing and inventory decisions. Our model allows us to offer an explanation from a supply chain perspective on the prevalence of trade credit practices as short-term financing sources for retailers. Our analysis reveals that optimally structured trade credit contracts have the supplier offering interest rates in the range $[0,\ r_f]$, with r_f in all likelihood the preferred rate as it gives incentives to the retailer to pay up front his order using his working capital to the extent possible. Thus, open account financing, with zero trade credit rate, is among the potentially optimal trade credit contracts. The supplier's equilibrium wholesale price depends on the retailer's "wealth" (working capital plus collateral): For rich retailers, it is in their Non-Bankruptcy Region; for poor ones, it is in their Bankruptcy Region; and for moderately rich ones, it can be either way, whichever gives the supplier a larger expected cash flow. To the supplier, the equilibrium in the retailer's Bankruptcy Region is preferred, as it leads to a larger expected cash flow by inducing a larger retailer's order quantity.

When offered an optimally structured supplier financing scheme, the retailer always prefers it to bank financing, and receives larger expected cash flows as long as he is not very poor. Unfortunately, for the low wealth retailers, since the expected cash flows they receive are lower under trade credit contracts, the alternative that the supplier offers a wholesale price contract without any financing provisions would have been preferable. But this is not an option the supplier will entertain offering as part of her Stackelberg equilibrium.

In summary, optimally priced trade credit is cheaper than bank credit, and improves supply chain efficiency by inducing larger retailer's order quantities. However, it does not perfectly coordinate the chain. It always leads to increased profitability for the supplier, and it also benefits most retailers (all but the "very poor" ones wealthwise). Thus, theoretically, the supplier can extend credits to any retailers at low interest rates (even if she might need bank financing herself).

REFERENCES

Ang J.S., Chua J.H., McConnell J.J. 1982. The administrative costs of corporate bankruptcy: A note. *The Journal of Finance*, 37(1):219–226.

Barnea A., Haugen R.A., Senbet L.W. 1981. Market imperfections agency problems, and capital structure: A review. *Financial Management*, 10(3):7–22.

Barro R.J. 1976. The loan market, collateral, and rates of interest, *Journal of Money, Credit and Banking*, 8(4):439–456.

Besanko D., Thakor A.V. 1987. Collateral and rationing: Sorting equilibria in monopolistic and competitive credit markets, *International Economic Review*, 28(3): 671–689.

Bester H. 1994. The role of collateral in a model of debt renegotiation. *Journal of Money, Credit and Banking*, 26(1):72–86.

Brealey R.A., Myers S.C., Marcus A.J. 1995. *Fundamentals of Corporate Finance*. McGraw-Hill.

Burkart M., Ellingsen T. 2004. In-kind finance: A theory of trade credit. *The American Economic Review*, 94(3):569–590.

Fabbri D., Klapper L. October 2008. Market power and the matching of trade credit terms. Policy Research Working Paper Series, The World Bank, Development Research Group, Finance and Private Sector Team. Available at: http://ideas.repec.org/p/wbk/wbrwps/4754.html.

Fisman R., Love I. 2003. Trade credit, financial intermediary development and industry growth. *Journal of Finance*, 58:353–373.

Giannetti M., Burkart M., Ellingsen T. 2007. What you sell is what you lend? Explaining trade credit contracts. Working paper. Stockholm Sweden: Stockholm School of Economics.

Gupta D. 2008. Technical note: Financing the Newsvendor. Working paper. University of Minnesota, Twin Cities.

Horen N.V. 2007. Customer market power and the provision of trade credit: Evidence from Eastern Europe and Central Asia. The World Bank Development Prospects Group.

Kouvelis P., Zhao W. 2008. Financing the newsvendor: Supplier vs. bank, and the structure of optimal trade credit contracts. Working paper. St. Louis, Missouri: Olin Business School, Washington University.

Kouvelis P., Zhao W. 2011. The newsvendor problem and price-only contract when bankruptcy costs exist. *Productions and Operations Management*, doi: 10.1111/j.1937-5956.2010.01211.x

Koziol C. 2007. Do good or bad borrowers pledge more collateral? *International Journal of Managerial Finance*, 3(2):132–163.

Lai G., Debo L.G., Sycara K. 2009. Sharing inventory risk in supply chain: The implication of financial constraint. *Omega*, 37(4):811–825.

Lariviere M.A., Porteus E.L. 2001. Selling to the newsvendor: An analysis of price-only contracts. *Manufacturing Service & Operations Management*, 3(4):293–305.

Marotta G. 2005. Is trade credit more expensive than bank credit loans? Evidence from Italian firm-level data. *Applied Economics*, 37:403–416.

Modigliani F., Miller M. 1958. The Cost of Capital, Corporation Finance and the Theory of Investment. *American Economic Review*, 48(3):261–297.

Ng C., Smith J., Smith R. 1999. Evidence on the determinants of credit terms used in interfirm trade. *Journal of Finance*, 54(3):1109–1129.

Petersen M.A., Rajan R. 1997. Trade credit, theory and evidence. *Review of Financial Studies*, 10:661–691.

Tirole J. 2006. *The theory of corporate finance.* Princeton, New Jersey: Princeton University Press.

Vandenberg P. 2003. Adapting to the financial landscape: Evidence from small firms in Nairobi. *World Development*, 31(11):1829–1843.

Xu X., Birge J.R. 2004. Joint production and financing decisions: Modeling and analysis. Working paper. Chicago, Illinois: The University of Chicago Graduate School of Business.

Zhou, J., Groenevelt H. 2007. Impacts of Financial Collaboration in a Three-Party Supply Chain. Working paper. Rochester, New York: The Simon School, University of Rochester.

CHAPTER ELEVEN

The Role of Financial Services in Procurement Contracts

RENÉ CALDENTEY AND XIANGFENG CHEN

In this chapter we investigate the interplay between operational and financial decisions within the context of a two-echelon supply chain. A retailer purchases a single product from a supplier and afterwards sells it in the retail market to a random demand. The retailer, however, is budget-constrained and is therefore limited in the number of units that he may purchase from the supplier. We study two alternative forms of financing that the retailer can use to overcome the limitations imposed by the budget constraint. First, we consider the case of *internal financing* in which the supplier offers financial services to the retailer in the form of a procurement contract. The type of contracts that we consider allows the retailer to pay *in arrears* a fraction of the procurement cost after demand is realized. Second, we consider the case of *external financing* in which a third party financial institution (e.g., a bank) offers a commercial loan to the retailer. Our results show that the performance of the entire supply chain can be severely

Acknowledgements: The authors are grateful to Lingxiu Dong and two anonymous reviewers for their valuable comments, which help improved the chapter's content and presentation. The research of the first author was supported in part by FONDECYT grant No. 1070342. The second author acknowledges support from NSF of China (No.70972046) and Shanghai Pujiang Program in 2009.

Handbook of Integrated Risk Management in Global Supply Chains, First Edition. Edited by Panos Kouvelis, Lingxiu Dong, Onur Boyabatli, and Rong Li.

affected by the lack of financing and that it is in the interest of both players to find ways to finance the retailer's operations. We also show that, for the most part, both the supplier and the retailer (and the entire supply chain) are better off by using *internal financing* rather than by relying on *external financing*. From the supplier's point of view, we show that the value of offering internal financing decreases with the size of the retailer's initial budget. This is despite of the fact that the risk of the retailer defaulting on his contractual obligations decreases with his initial budget. Interestingly, from the retailer's standpoint the value of internal financing is non-monotonic in his budget and there is an intermediate value at which his expected payoff is maximized.

11.1 Introduction

A key function of supply chain management is the effective coordination of material flows, information flows, and financial flows across different organizations. For the most part, the research in operations management has focused on optimizing the interplay between material flows and information flows ignoring the issue of coordinating material and financial flows. In practice, however, companies have limited working capital to operate and financial constraints play a pivotal role in determining firms' production and procurement decisions. In the global supply chain, and especially during financial crisis and economic downturns, these financial constraints are strengthened by common cash-management practices that promote collecting account receivables as quick as possible while postponing payments to providers and suppliers. This is not only true for small start-up companies with limited access to credit but also for large corporations. However, this "war for cash" (see Milne 2009) puts extra pressure on small companies that usually find themselves making suboptimal procurement decisions to balance their cash reserves increasing their credit risk and their chances of getting out of business.

In this chapter we investigate the connection between operational and financial decisions within the context of a two-echelon supply chain. We consider a stylized model in which a retailer purchases a single product from a supplier and afterwards sells it in the retail market to a random demand. The retailer has a fixed budget that limits his procurement decisions. In this setting, we propose and analyze two alternative forms of financing that the retailer can use to overcome the limitations imposed by the budget constraint. First, we consider the case of what we call *internal financing* in which the supplier offers financial services to the retailer in the form of a procurement contract. The type of *credit contracts* that we consider allow the retailer to pay *in arrears* a fraction of the procurement cost after demand is realized. By offering this option to delay payments, the supplier is effectively allowing the retailer to order more so as to take advantage of a possible upside on market demand. On the downside, the supplier is also internalizing some of the retailer's demand risk. The second form of financing that we consider is what we call *external financing* in which a third party financial institution (e.g., a bank) offers a commercial loan to the retailer. In this case, the supplier offers a traditional wholesale contract in which the retailer must pay in advance (possibly using the bank's loan) the totality of the procurements costs. In contrast to the credit contract, in this case it is the bank, and not the supplier, who is getting exposed to the retailer's credit risk.

The main goal of this chapter is to shed some light on how financial services impact agents' operational and financial decisions and how they should be designed and used to create value in a supply chain environment. As their names suggest, our choice of these two contrasting forms of financing, *internal* and *external*, is intended to highlight how this process of value creation is affected by the level of involvement that the agent providing the financing has within the supply chain. For instance, in the case of internal financing, it is in the supplier best interest to support the retailer's business as this will have a direct impact on the supplier's own business. On the other hand, a bank providing external financing has no formal stakes in the supply chain and so should be less concern with its operations as long as the loan and interests are repaid.

Our particular interest in the internal financing model is motivated by some successful applications in real business operations. As an example, consider the case of Chinese Material Shortage Transportation Group (CMST), one of the largest logistics enterprises in China. It is common for many small- and medium-sized paper manufacturers in mainland China to purchase materials from international suppliers. While most of these companies have limited working capital, it is still the case that the financial system is unable to provide adequate services to support their businesses. As a result, most of these small manufacturers find themselves making suboptimal procurement decisions. CMST viewed this gap as a business opportunity and, since 2002, started financing these small paper manufacturers to buy paper materials from international suppliers while simultaneously providing the logistics services required in these transactions. The financial service offered by CMST is essentially a credit contract under which the paper manufacturer pays a fraction of the wholesale price charged by the supplier as a deposit and CMST covers the difference. Then, the manufacturer repays CMST the remaining fraction of the wholesale price when the sale season is over. With this supply chain financing service, CMST has become one of the leading logistics-financing service providers in China. In 2006, CMST financing supply chain business was about 1.1 billion dollars, up from 750 million dollars in 2005. Currently, this operations is one of most popular modes of financing for small- and medium-sized companies in China.

As the CMST example unveils, bringing financial services into supply chain management has the potential to improve the operational efficiency and the profits of the entire supply chain. It is because of this value-creation capacity that we expect financial services to play an even more predominant role in the future growth of global supply chain management and it is precisely this fact that motivates this research.

We conclude the introduction by attempting to position our chapter within—but not reviewing—the vast literature on supply chain management. We refer the reader to the books by de Kok and Graves (2003) and Simchi-Levi et al. (2004)

for a general overview of supply chain management issues and to the survey article by Cachon (2003) for a review of supply chain management contracts.

One of the distinguishing features of our model with respect to most of the existing literature in supply chain management is the consideration of a budget constraint that limits the retailer's procurement capacity. A few recent exceptions are the papers by Buzacott and Zhang (2004), Caldentey and Haugh (2009), Dada and Hu (2008), Kouvelis and Zhao (2008), and Xu and Birge (2004).

Xu and Birge (2004) analyze a single-period newsvendor model, which is used to illustrate how a firm's inventory decisions are affected by the existence of a budget constraint and the firm's capital structure (debt/equity ratio). Hu and Sobel (2005) use a multiperiod production model to examine the interdependence of a firm's capital structure and its short-term operating decisions concerning inventory, dividends, and liquidity. In a similar setting, Dada and Hu (2008) consider a budget-constrained newsvendor that can borrow from a bank that acts strategically when choosing the terms (interest rate) of the loan. The paper characterizes the Stackelberg equilibrium and investigates conditions under which channel coordination (i.e., the budget-constrained newsvendor orders the same amount than the unconstrained newsvendor) can be achieved.

Buzacott and Zhang (2004) consider a deterministic multiperiod production/inventory control model and investigate the interplay between inventory decisions and asset-based financing. In their model, a retailer finances its operations by borrowing from a commercial bank. The terms of the loans are structured contingent upon the retailer's balance sheets and income statements (in particular, inventories and receivables). The authors conclude that asset-based financing allows retailers to enhance their cash return over what it would be if they were only able to use their own capital.

The work by Caldentey and Haugh (2009) and Kouvelis and Zhao (2008) are the most closely related to this chapter. They both consider a two-echelon supply chain system in which the retailer is budget constrained and investigate different types of procurement contracts between the agents using a Stackelberg equilibrium concept. In Caldentey and Haugh (2009), the supplier offers a menu of wholesale contracts (with different execution times and wholesale prices) and the retailer choose the optimal timing at which to execute the contract. The reason why the timing of the contract is important is because in their model the retailer's demand is partially correlated to a financial index that the retailer (and the supplier) can track. As a result, the retailer can dynamically trade in the financial market to adjust his budget to make it contingent upon the evolution of the index and choose the appropriate time at which to execute the contract with the supplier. Their model shows how financial markets can be used as a source of public information upon which procurement contracts can be written and as a means for financial hedging to mitigate the effects of the budget constraint.

Similar to our work, in Kouvelis and Zhao (2008) the supplier takes a proactive role in offering different type of contracts designed to provide financial services to the budget-constrained retailer. The authors analyze a set of alternative financing schemes including *supplier early payment discount*, *open account financing*, *joint supplier financing with bank*, and *bank financing*. They conclude that in an optimally designed scheme it is in the supplier's best interest to offer financing to the retailer and the retailer will always prefer financing from the supplier rather than the bank. We reach similar conclusions in our setting. A noticeable difference between their formulation and ours is that in our model we impose a budget constraint at the time the contract is signed that explicitly limits the retailer's ordering decision. With the possibility of the retailer declaring bankruptcy, this constraint limits the supplier's default risk exposure.

Finally, it is worth mentioning that there exits a somehow related stream of research that investigates the use of financial markets and financial instruments (such as forwards and futures contracts, options, swaps, etc.) to hedge operational risk exposure (see Boyabatli and Toktay (2004) for a detailed review). For example, Caldentey and Haugh (2006) consider the general problem of dynamically hedging the profits of a risk-averse corporation when these profits are partially correlated with returns in the financial markets. Gaur and Seshadri (2005) consider a risk-averse newsvendor model in which demand is perfectly and partially correlated with a marketable risky security. In both cases, they show how the retailer can hedge demand uncertainty by trading on the risky security. Chod et al. (2009) examine the joint impact of operational flexibility and financial hedging on a firm's performance and their complementary/substitutability with the firm's overall risk management strategy. Ding et al. (2007) and Dong et al. (2006) examine the interaction of operational and financial decisions from an integrated risk management standpoint. Boyabatli and Toktay (2010) analyze the effect of capital market imperfections on a firm's operational and financial decisions in a capacity investment setting. The authors consider that the firm can use tradeable asset's forward contracts and commercial loans to relax its budget constraint. Babich and Sobel (2004) propose an infinite-horizon discounted Markov decision process in which an IPO event is treated as a stopping time. The value of the IPO is modeled as a random variable whose distribution depends on the firm's current assets, its most recent sales revenue, and its most recent profits. Every period the firm must decide on capacity expansion, production, and loan size. They characterize an optimal capacity-expansion and financing policy so as to maximize the expected present value of the firm's IPO. Babich et al. (2008) study how trade credit financing affect the relationship among firms in supply chain, supplier selection, and supply chain performance.

The rest of this chapter is organized as follows. In the next section we present the mathematical formulation including the retailer and supplier payoff functions, the main features of the credit contract and some fundamental assumptions about the market demand. We also characterize the Stackelberg equilibrium under a traditional wholesale contract which we use as a benchmark for comparison throughout the chapter. In Section 11.4 we investigate the optimal design of the credit contract using a non-cooperative game theoretical framework. First, we compute the retailer best response ordering strategy as a function of the parameters of the credit contract. Then, we solve the supplier's optimization problem to determine the optimal parameters of this contract. We conclude this section analyzing in detail the extreme case in which the retailer has no initial budget. Section 11.5 is devoted to the study of the external financing model. We first define the notion of a feasible loan in a competitive financial market environment. We then use this concept to solve for the optimal loan. We show that the outcome of the game in this case is equivalent to a model in which the retailer has an unlimited budget (a reminiscence of the Modigliani and Miller's irrelevance principle). Section 11.6 presents a set of numerical experiments that we use to highlight the main features and insights of our model. In these experiments, we compare the outcome of the internal and external financing models from the point of view of the agents' payoffs as well as the entire supply chain efficiency. We also investigate the impact of demand variability on the outcome of the game and the choice of the best contract. Concluding remarks and possible extensions of our model are discussed in Section 11.7. In particular, we emphasize the extension that considers the case in which the retailer's initial budget is private knowledge and propose some simple variations of the credit contract that could be used to partially solve the supplier's adverse selection problem.

11.2 Model Description

We model an isolated portion of a competitive supply chain with one supplier that produces a single product and one retailer that faces a random market demand D that is realized at a future time T.[1] We assume that D is a non-negative random variable with distribution function $F(D)$. We will make the following assumptions about F throughout the chapter.

Assumption 11.1 *The demand distribution function F satisfies the following properties:*

(i) It has a smooth density $f(D) > 0$ *in* (a, b), *for* $0 \leq a \leq b \leq \infty$,
(ii) It has a finite mean, and
(iii) Its hazard function $h(D) := f(D)/\bar{F}(D)$ *is increasing in* $D \geq 0$, *where* $\bar{F}(D)$ *is the tail distribution* $\bar{F}(D) := 1 - F(D)$.

These are not particularly restrictive requirements on $F(D)$[2] that we impose to guarantee the existence and uniqueness of an equilibrium (see Lemma 11.1 below).

A distinctive feature of our model is that the retailer is restricted by a budget constraint that limits his ordering decisions. In particular, we assume that the retailer has an initial budget B that may be used to purchase product units from the supplier. On the other hand, we assume that the supplier has *deep pockets*, that

[1] Similar models are discussed in detail in Section 2 of Cachon (2003). See also Lariviere and Porteus (2001).
[2] They are satisfied by most popular distributions such as the Uniform, Log-Normal, Exponential and Weibull, among many others (see Lariviere 2006).

is, enough working capital to pay for manufacturing costs independent of the size of the production batch.

At time $t = 0$, the retailer and supplier negotiate the terms of a procurement contract that specifies the following quantities:

- **Credit Contract ($\boldsymbol{w}, \boldsymbol{\alpha}$)**: where w is the wholesale price per unit and $\alpha \in [0, 1]$ is the fraction of this wholesale price that the retailer must pay in advance at time $t = 0$. We will refer to α as the *credit parameter*.
- **Order Quantity $\boldsymbol{Q}$**: The number of units purchased by the retailer.

We assume that in the negotiation of this contract the supplier acts as a Stackelberg leader. That is, at $t = 0$ the supplier moves first and proposes the contract (w, α), to which the retailer then reacts by selecting the ordering level Q and paying the supplier the amount $\alpha\, w\, Q$. The remaining portion $(1 - \alpha)\, w\, Q$ is paid in arrears at time $t = T$, after the value of D is realized and the retailer's revenues are collected. These revenues are equal to $p \min\{D, Q\}$, where p is a fixed retail price. Neither a salvage value nor a return policy for unsold units are considered in this model. For notational convenience, we will normalize all prices in this economy so that $p = 1$. As a result, w, c, and B are measured relative to the retail price.

The implication of the retailer's limited budget on the execution of the contract is twofold. First, the order quantity placed at $t = 0$ must satisfy the *budget constraint* $\alpha\, w\, Q \leq B$. Second, if the realized demand D is sufficiently low then the retailer will be unable to pay the supplier the full amount $(1 - \alpha)\, w\, Q$ due at time T. In this case (which occurs if $\min\{D, Q\} + B < w\, Q$) the retailer declares bankruptcy and the supplier collects $\min\{D, Q\} + B - \alpha\, w\, Q$ instead of $(1 - \alpha)\, w\, Q$. Hence, it follows that the supplier faces a trade-off when selecting the optimal credit contract (w, α) to offer. On one hand, the supplier would like to choose a small α to minimize the impact of the budget constraint so as to boost the retailer's ordering level. Indeed, by choosing the contract (w, α) the supplier is effectively offering the retailer a loan of $(1 - \alpha)\, w\, Q$ to procure more units. On the other hand, the supplier would like to choose a large α to minimize the credit risk associated with the retailer defaulting at time T. We will discuss in full detail this trade-off in the following section.

In terms of *who knows what*, we consider for most part of this chapter the symmetric information case where all information is common knowledge. In particular, we assume that the supplier knows the retailer budget B, the demand probability distribution $F(D)$, and the retail price. We will discuss this symmetric information assumption in Section 11.7 where we discuss the case in which the retailer's budget B is private information.

For a given contract (w, α), we define the retailer's *net* expected payoff as a function of Q to be equal to $\pi_{\mathrm{I}}^{\mathrm{R}}(Q) := \mathbb{E}\big[\big(\min\{D, Q\} + B - w\, Q\big)^+ - B\big]$, where $\mathbb{E}[\cdot]$ denotes expectation with respect to F. (We use the subscript/superscript "I"—which stands for *internal* financing—to denote quantities related to the credit contract.) Note that in our definition of $\pi_{\mathrm{I}}^{\mathrm{R}}$ we have subtracted the initial budget B from the retailer's profits. Hence, $\pi_{\text{I}}^{\text{R}}$ measures the net contribution to earnings that the retailer gets by operating in this supply chain. For example, with this definition if the retailer chooses $Q = 0$ then his net payoff would be 0 reflecting the fact that he has gained nothing from his retail business. The retailer's optimal *net* expected payoff is obtained solving:

$$\Pi_{\text{I}}^{\text{R}} = \max_{Q \geq 0} \pi_{\text{I}}^{\text{R}}(Q) = \max_{Q \geq 0} \mathbb{E}\left[\left(\min\{D, Q\} + B - w\,Q\right)^{+}\right] - B \qquad (11.1)$$

$$\text{subject to} \quad \alpha\, w\, Q \leq B \qquad (11.2)$$

The positive part in the definition of Π^{R} captures the retailer's *limited liability* in case of bankruptcy.

Let us denote by Q_{I} the optimal solution to (11.1)–(11.2), which represents the retailer's *best response* to the supplier contract (w, α). Naturally, Q_{I} depends on w, α, and B. When we wish to emphasize this dependence we will include these quantities as part of its argument. For example, we will write $Q_{\text{I}}(w, \alpha)$ when discussing the dependence of the optimal ordering level on the terms of the contract (w, α). A similar convention applies to $\Pi_{\text{I}}^{\text{R}}$ and other quantities that we introduce below.

Under the common knowledge assumption, the supplier is able to anticipate the retailer's best response Q_{I}. As a result, the supplier chooses an optimal credit contract by solving

$$\Pi_{\text{I}}^{\text{S}} = \max_{w,\alpha} \mathbb{E}\left[(w - c)Q_{\text{I}}(w, \alpha) - \left(\min\{D, Q_{\text{I}}(w, \alpha)\} + B - wQ_{\text{I}}(w, \alpha)\right)^{-}\right] \qquad (11.3)$$

Note that c is the per unit manufacturing cost incurred by the supplier. Let us denote by $w_{\text{I}}(B)$ and $\alpha_{\text{I}}(B)$ the optimal solution to (11.3). Since we have assumed that the supplier is not budget constrained, $\Pi_{\text{I}}^{\text{S}}$ represents the supplier's operating profits, which can be negative.

To ensure the operability of the supply chain the market price must exceed the manufacturing cost, that is, $c \leq 1$. The difference $1 - c$ represents the net margin per unit sold made by the entire supply chain. This margin is split into $1 - w$ that goes to the retailer and $w - c$ that goes to the supplier. Naturally, we expect in equilibrium the supplier to set w so that $c < w \leq p$ creating the so-called *double marginalization* inefficiency (e.g., Spengler 1950). We will measure this inefficiency by computing the *competition penalty*, which is defined as one minus the decentralized supply chain payoff divided by the centralized supply chain payoff (see Cachon and Zipkin 1999). That is:

$$\mathcal{P} := 1 - \frac{\Pi^{\text{R}} + \Pi^{\text{S}}}{\Pi^{\text{C}}} \qquad (11.4)$$

In (11.4) Π^{C} is the optimal centralized payoff obtained by a central planner that owns both the supply and retail operations. (We will use a subscript/superscript "C" to denote quantities related to this centralized solution.) In order to have a fair comparison between the centralized and decentralized operations, we need to make certain assumptions about Π^{C}. In particular, we will assume that the central planner, like the supplier, has enough working capital to pay for the manufacturing costs. Hence, using this centralized solution as a benchmark provides a measure to compute the benefits of vertical integration. It follows that:

$$\Pi^{C} = \max_{Q \geq 0} \mathbb{E}\Big[\min\{D, Q\} - cQ \Big] \tag{11.5}$$

The optimal centralized inventory level is $Q_C = \bar{F}^{-1}(c)$. The following are some additional remarks about the model.

1. The credit parameter, α, allows to transfer part of the demand risk from the retailer to the supplier. Indeed, if $\alpha = 0$ then the supplier is effectively offering the retailer the option to buy as many units as he wants and to pay for them in arrears after demand is realized. This can be a risky strategy for the supplier, especially if the retailer overestimates market demand. On the other hand, when $\alpha = 1$ the supplier faces no risk since all payments are made at time $t = 0$.
2. In our formulation of Π^{R} we have implicitly assumed that the retailer has no other investment opportunity besides his retail operations. As a result, his excess wealth $B - \alpha\, w\, Q$ at time $t = 0$ generates no interest at time $t = T$. Alternatively, we can think that the retailer has access to a risk-free cash account with zero interest rate $r_f = 0$. The reason why we restrict the retailer's investment opportunities is to isolate the role that financial supply chain management (represented here by our credit contract) plays as a value generating activity. We refer the reader to Caldentey and Haugh (2009) for a model that explicitly considers the existence of financial markets as an alternative investment opportunity for the retailer.
3. It is worth noticing that the *traditional wholesale contract* (e.g., see Cachon 2003), in which the retailer must pay the full procurement cost $w\,Q$ at time $t = 0$, is a special case of our credit contract with $\alpha = 1$. Because of the popularity of the wholesale contract in both research and practice, we will use it throughout this chapter as a benchmark for comparisons.

11.2.1 SUMMARY OF NOTATION

For future references, and to help the reader keep track of the different components of our model, let us summarize here some of the notation that we will use throughout the chapter. Additional notation will be introduced later on as needed.

- $H(Q) := Q\,h(Q)$, the *generalized failure rate* function of the demand D
- $\hat{w} := \text{argmax}\{w\,\bar{F}^{-1}(w) \,:\, w \in [0, 1]\}$
- $w_E := \text{argmax}\{(w - c)\,\bar{F}^{-1}(w) \,:\, w \in [c, 1]\}$
- $\bar{w}(B) := \max\{w \in [0, 1] \,:\, w\,\bar{F}^{-1}(w) \geq B\}$ if $B \leq \hat{w}\,\bar{F}^{-1}(\hat{w})$
- $\underline{w}(B) := \min\{w \in [0, 1] \,:\, w\,\bar{F}^{-1}(w) \geq B\}$ if $B \leq \hat{w}\,\bar{F}^{-1}(\hat{w})$

- $\hat{Q} := \bar{F}^{-1}(\hat{w})$
- $\tilde{Q}(w, B)$ solves $\bar{F}(\tilde{Q}) = w\,\bar{F}(w\,\tilde{Q} - B)$
- $Q_{\mathrm{C}} := \bar{F}^{-1}(c)$
- $Q_{\mathrm{E}} := \bar{F}^{-1}(w_{\mathrm{E}})$
- $\widehat{B} := \hat{w}\,\hat{Q}$
- $B_{\mathrm{E}} := w_{\mathrm{E}}\,Q_{\mathrm{E}}$

It follows from Assumption 11.1 that function $w\,\bar{F}^{-1}(w)$ and $(w - c)\,\bar{F}^{-1}(w)$ are unimodal. As a result, $w\,\bar{F}^{-1}(w) \geq B$ for all $w \in [\underline{w}(B), \bar{w}(B)]$. Also, one can show that $\hat{Q}$ solves $H(\hat{Q}) = 1$. Such a solution exists and is unique since, by Assumption 11.1, $H(Q)$ is an increasing function such that $H(0) = 0$ and $\lim_{Q\to\infty} H(Q) > 1$ (see Theorem 2 in Lariviere [2006]). The existence and uniqueness of $\tilde{Q}(w, B)$ are discussed below in Lemma 11.1.

11.3 Wholesale Contract with a Budget Constraint (w_{T}, Q_{T})

Let us conclude this section reviewing the Stackelberg-Nash equilibrium for the *traditional* wholesale for the case in which the retailer is budget constrained. We will use a subscript/superscript "T" to denote quantities related to the market equilibrium under this contract such as the wholesale price $w_{\mathrm{T}}(B)$, the retailer's ordering level $Q_{\mathrm{T}}(B)$, and the agents' expected payoff $\Pi^{\mathrm{R}}_{\mathrm{T}}(B)$ and $\Pi^{\mathrm{S}}_{\mathrm{T}}(B)$, as a function of the retailer's budget B.

To compute this equilibrium, we first solve the retailer's optimization problem which is given by:

$$\max_{0\leq Q\leq B/w} \mathbb{E}\left[\min\{D, Q\} - w\,Q\right]$$

This is a newsboy problem and the optimal ordering level is given by $Q^*(w, B) = \min\{B/w, \bar{F}^{-1}(w)\}$. It follows from the definition of $\bar{w}(B)$ and $\underline{w}(B)$ in the previous section that $Q^*(w, B) = B/w$ if and only if $w \in [\underline{w}(B), \bar{w}(B)]$.

Based on the retailer's optimal ordering level, the supplier chooses the wholesale price that maximizes his payoff, that is:

$$w_{\mathrm{T}}(B) = \underset{c\leq w\leq 1}{\mathrm{argmax}}\{(w - c)\,Q^*(w, B)\}\}.$$

Using the results in Lariviere and Porteus (2001) we can show that the first-order optimality condition is also sufficient when F has *increasing generalized failure rate* (IGFR). In our case, Assumption 11.1 ensures that F has IGFR. An alternative characterization of the wholesale price $w_{\mathrm{T}}(B)$ is given by $w_{\mathrm{T}}(B) = \max\{w_{\mathrm{E}}, \bar{w}(B)\}$.

The ordering quantity is given by:

$$Q_T(B) = \min\left\{\frac{B}{w_T(B)}, \bar{F}^{-1}(w_T(B))\right\}$$

The retailer and supplier equilibrium payoffs are:

$$\Pi^R_T(B) = \mathbb{E}\left[\min\{D, Q_T(B)\} - w_T(B)\, Q_T(B)\right]$$
$$\text{and} \quad \Pi^S_T(B) = (w_T(B) - c)\, Q_T(B).$$

The following result summarizes some useful properties of the equilibrium under the traditional wholesale contract as a function of B. The proof is straightforward and it is omitted.

Proposition 11.1 *The wholesale price $w_T(B)$ is decreasing in B while $Q_T(B)$, $\Pi^R_T(B)$ and $\Pi^R_T(B)$ are all increasing in B. In particular, the wholesale price and ordering level satisfy:*

$$(w_T(B), Q_T(B)) = \begin{cases} (1, 0) & \text{if} \quad B = 0 \\ (w_E, Q_E) & \text{if} \quad B \geq B_E \end{cases}$$

According to this result, the budget B_E is the minimum budget that the retailer needs to purchase the optimal unconstrained quantity Q_E. The budget B_E will be useful to formalize the notion of a *large budget* (see Proposition 11.5). The subscript/superscript "E" stands for *external* financing, and the reason for this choice is that (as we will see in Section 11.5) w_E and Q_E are the equilibrium wholesale price and production level when the retailer uses a third party financial institution to finance his operations instead of the credit contract.

11.4 Equilibrium Under a Credit Contract (Q_I, w_I, α_I)

In this section, we characterize the equilibrium for a supply chain that operates under the credit contract described in the previous section. As it is customary when determining a Stackelberg-Nash equilibrium, we first compute the *follower's* best response as a function of an arbitrary strategy selected by the *leader*. That is, we start solving the retailer's optimization problem in (11.1)–(11.2) to find $Q_I(w, \alpha)$ for a fixed contract (w, α). Then, we will plug this solution into the supplier's optimization in (11.3) to compute the optimal contract (w_I, α_I).

11.4.1 RETAILER'S OPTIMAL ORDERING STRATEGY (Q_I)

In the process of computing the equilibrium we would like to be able to identify the main differences between our solution and existing results on procurement contracts. In particular, we would like to understand the impact that both the retailer's budget constraint and limited liability have on Q_I. For this, we find convenient to isolate these two effects by considering the following optimization problem:

$$\max_{Q \geq 0} \mathbb{E}\left[\left(\min\{D, Q\} + B - w\,Q\right)^{+} - B\right] \tag{11.6}$$

This corresponds to the retailer's original problem without the budget constraint $\alpha\, w\, Q \leq B$. Let $\tilde{Q}$ be the solution to this unrestricted problem. Note that we can interpret (11.6) as the retailer's problem if the supplier offers the contract $(w, 0)$ with full financing at time 0, and so $\tilde{Q}(w, B) = Q_I(w, 0, B)$. The following is a useful intermediate step in our characterization of Q_I.

LEMMA 11.1. *Suppose F satisfies the conditions in Assumption 11.1. Then, if $w < 1$ there exits a unique non-negative solution $\tilde{Q}(w, B)$ to (11.6) that solves:*

$$\bar{F}(\tilde{Q}) = w\,\bar{F}(w\,\tilde{Q} - B)$$

$\tilde{Q}(w, B)$ is a decreasing function of both w and B, and satisfies:

$$\tilde{Q}(w, B) = \bar{F}^{-1}(w), \quad \textit{for all } B \geq w\,\bar{F}^{-1}(w)$$

The function $w\,\tilde{Q}(w, B)$ is unimodal in $w \in [c, 1]$ and attains its maximum at w_0 such that $\tilde{Q}(w_0, B) = \hat{Q}$.

Proof: See the Appendix at the end of the chapter.

The reason to introduce $\tilde{Q}$ is twofold. On one hand, it will be useful in our characterization of Q_I in Proposition 11.2. Second, it represents the optimal inventory level for a retailer that has exclusively limited liability and no budget constraint when ordering at $t = 0$.

To get some intuition about the value of $\tilde{Q}(w, B)$, recall that in the traditional newsvendor model (e.g., Hadley and Whitin (1963)), with full liability, the retailer's optimal solution solves the familiar fractile equation $\bar{F}(Q) = w$ (or equivalently, $Q = \bar{F}^{-1}(w)$). This is a first-order optimality condition that requires the marginal revenue of an extra unit, $\bar{F}(Q)$, to be equal to the marginal cost of this extra unit, w. In our case, the value of $\tilde{Q}$ solves a modified fractile equation $\bar{F}(\tilde{Q}) = w\,\bar{F}(w\,\tilde{Q} - B)$. Since $w \geq w\,\bar{F}(w\tilde{Q} - B)$, it follows that a retailer with limited liability faces lower marginal costs and so $\tilde{Q} \geq \bar{F}^{-1}(w)$. This is not particularly surprising if we interpret this limited liability as an option that reduces the retailer's downside risk associated with low demand outcomes. Interestingly, this marginal cost $w\,\bar{F}(w\,\tilde{Q} - B)$ increases with the retailer's budget B. As a result, $\tilde{Q}$ decreases with B. Hence, for B sufficiently large $Q_I = \tilde{Q}$ and so the retailer optimal ordering level decreases with his initial budget. Intuitively, this happens because retailers with smaller budgets have less stakes at risk when choosing their inventory levels (as their potential loses are bounded by B) and so they order more aggressively to take full advantage of high demand outcomes.

Figure 11.1 depicts the retailer net payoff as a function of D for a fixed order Q and three different budgets $0 = B_0 < B_1 < B_2$.

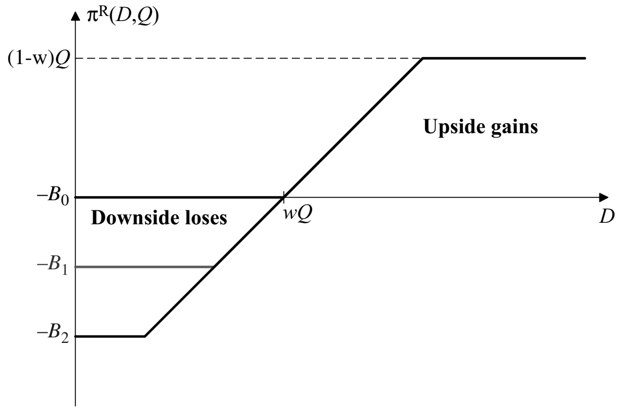


FIGURE 11.1 **Retailer's net payoff** $\pi_I^R(D, Q) = \left(\min\{D, Q\} + B - w\,Q\right)^+ - B$ **as a function of** D **for a fixed order** Q **and three different budgets** $0 = B_0 < B_1 < B_2$.

As we can see, the retailer's downside loses increase with B while his upside gains remain constant, all else being equal. This asymmetry in the retailer's payoff explains why small-budget retailers have the incentives to order more than their high-budget counterparts so as to take advantage of high demand scenarios.

Based on Lemma 11.1 we derive in the following proposition the optimal ordering level for the retailer.

Proposition 11.2 *Suppose the retailer has a budget B and the supplier chooses the contract* (w, α). *Then, under the assumptions on Lemma 11.1 the retailer's optimal solution* Q_I *satisfies:*

$$Q_I(w, \alpha, B) = \min\left\{\tilde{Q}(w, B)\,,\ \frac{B}{\alpha\, w}\right\}$$

Proof: See the Appendix at the end of the chapter.

Table 11.1 provides a summary of the possible values of Q_I as well as a comparison with $\bar{F}^{-1}(w)$ as a function of B. (Recall that $w\,\bar{F}^{-1}(w)$ represents the minimum budget required to finance the optimal noncooperative solution $\bar{F}^{-1}(w)$ under a wholesale contract with no budget constraint.)

TABLE 11.1 **Retailer's Optimal Ordering Level as a Function of** ***B***

Budget Range		Q_I	Q_I v.s. $\bar{F}^{-1}(w)$
Small	$B \le \alpha\, w\, \bar{F}^{-1}(w)$	$B/(\alpha\, w)$	$Q_I \le \bar{F}^{-1}(w)$
Small to Medium	$\alpha\, w\, \bar{F}^{-1}(w) \le B \le \alpha\, w\, \tilde{Q}(w)$	$B/(\alpha\, w)$	$Q_I \ge \bar{F}^{-1}(w)$
Medium to Large	$\alpha\, w\, \tilde{Q}(w) \le B \le w\, \bar{F}^{-1}(w)$	$\tilde{Q}(w)$	$Q_I \ge \bar{F}^{-1}(w)$
Large	$w\, \bar{F}^{-1}(w) \le B$	$\bar{F}^{-1}(w)$	$Q_I = \bar{F}^{-1}(w)$

It follows from Table 11.1 that a budget constrained retailer orders less than his unconstrained counterpart only if $B \le \alpha\, w\, \bar{F}^{-1}(w)$. However, we will show below (in Corollary 11.1) that this is never the case under an optimal contract $(w_{\rm I}, \alpha_{\rm I})$. Also, according to Table 11.1, $w\, \bar{F}^{-1}(w)$ is an upper bound on how much a retailer is willing to pay the supplier at $t = 0$, that is, $\alpha\, w\, Q_{\rm I}(w,\alpha) \le w\, \bar{F}^{-1}(w)$ for all contract (w, α). However, this condition does not rule out that $w\, Q_{\rm I}(w,\alpha) > w\, \bar{F}^{-1}(w)$. Indeed, we will show that in equilibrium this strict inequality does hold when B is relatively small.

Table 11.1 was built for a fixed contract (w, α). In the following section we characterize the supplier's optimal credit contract $(w_{\rm I}, \alpha_{\rm I})$ and show that at optimality $\alpha_{\rm I}\, w_{\rm I}\, \bar{F}^{-1}(w_{\rm I}) \ge B$. Hence, it is in the supplier best interest to select a contract that induces the retailer to take advantage of the credit line offered by the supplier.

11.4.2 SUPPLIER'S OPTIMAL CONTRACT $(w_{\rm I}, \alpha_{\rm I})$

It follows from equation (11.3) that the supplier's optimal contract solves:

$$\begin{aligned}\Pi_{\rm I}^{\rm S} &= \max_{w\in[c,1],\,\alpha\in[0,1]} \mathbb{E}\left[(w-c)\,Q_{\rm I}(w,\alpha) - \Big(D + B - w\,Q_{\rm I}(w,\alpha)\Big)^{-}\right] \\ &= \max_{w\in[c,1],\,\alpha\in[0,1]} \left\{(w-c)\,Q_{\rm I}(w,\alpha) - \int_0^{(w\,Q_{\rm I}(w,\alpha)-B)^+} F(D)\,\mathrm{d}D\right\}\end{aligned}$$

We compute $(w_{\rm I}, \alpha_{\rm I})$ in two steps. First, we determine the optimal value $\alpha_{\rm I}(w)$ for a fixed value of w. Then, we compute the optimal wholesale price $w_{\rm I}$.

According to Proposition 11.2, the retailer's optimal ordering decision $Q_{\rm I}(w,\alpha)$ is given by:

$$Q_{\rm I}(w,\alpha) = \min\left\{\tilde{Q}(w)\,,\ \frac{B}{\alpha\, w}\right\}, \quad \text{where } \tilde{Q}(w) \text{ solves } \ \bar{F}(\tilde{Q}) = w\,\bar{F}(w\,\tilde{Q} - B)$$

For a fixed w, let us define $\tilde{\alpha}(w, B) := B/(w\,\tilde{Q}(w))$. It follows that if $\alpha \le \tilde{\alpha}(w, B)$ then $Q_{\rm I}(w,\alpha) = \tilde{Q}(w)$, which is independent of α. Hence, without a significant loss in generality, we can assume that at optimality $\alpha_{\rm I}(w) \ge \min\{1, \tilde{\alpha}\}$.[3] In particular, when B is large (i.e., $B \ge w\,\bar{F}^{-1}(w)$), $\tilde{\alpha} \ge 1$ and so $\alpha_{\rm I}(w) = 1$. That is, in this case the supplier does not need to offer financial support to the retailer. On the other hand, for $B \le w\,\bar{F}^{-1}(w)$ we have $\tilde{\alpha} \le 1$ and we can restrict the search for an optimal $\alpha_{\rm I}(w)$ to the domain $[\tilde{\alpha}(w), 1]$. In this range $Q_{\rm I}(w,\alpha) = B/(\alpha\, w)$ and the supplier's problem reduces to:

$$\Pi_{\rm I}^{\rm S} = \max_{\alpha\in[\tilde{\alpha}(w),1]} \left(1 - \frac{c}{w}\right)\frac{B}{\alpha} - \mathbb{E}\left[\left(D - (1-\alpha)\,\frac{B}{\alpha}\right)^{-}\right], \quad (B \le w\,\bar{F}^{-1}(w))$$

[3] We assume that if the supplier is indifferent between α_1 and α_2 then he always selects $\max\{\alpha_1, \alpha_2\}$ so as to maximize the payment he receives at time 0.

Proposition 11.3 *Suppose F satisfies the conditions in Assumption 11.1. Then, for a fixed w such that* $c < w < 1$ *the optimal* $\alpha_{\rm I}(w, B)$ *is given by:*

$$\alpha_{\rm I}(w, B) = \max\left\{\frac{B}{B + \bar{F}^{-1}\left(\frac{c}{w}\right)}, \frac{B}{B + \bar{F}^{-1}\left(\frac{1}{w}\bar{F}(\tilde{Q}(w))\right)}\right\} \in [0, 1]$$

In particular, $\alpha_{\rm I}(w, B) = 1$ *if and only if* $B \geq w\,\bar{F}^{-1}(w)$.

Proof: See the Appendix at the end of the chapter

It follows from Proposition 11.3—and the fact that $\tilde{Q}(w)$ decreases with B—that $\alpha_{\rm I}(w, B)$ is monotonically increasing in the retailer's budget. When $B = 0$, the supplier offers the retailer 100% financing by setting $\alpha_{\rm I}(w, 0) = 0$. (Of course, this is the only choice of α that allows the retailer and the entire supply chain to operate.) On the other hand, when the budget is large (i.e., $B \geq w\,\bar{F}^{-1}(w)$), the retailer does not need any financial support as he is able to pay the entire procurement cost at $t = 0$. As a result, the supplier sets $\alpha_{\rm I}(w, B) = 1$ for $B \geq w\,\bar{F}^{-1}(w)$.

As a byproduct of Proposition 11.3, we can obtain a simple lower bound for $\alpha_{\rm I}(w, B)$ (uniform on w). From the first term inside the maximum it follows that:

$$\alpha_{\rm I}(w, B) \geq \frac{B}{B + \bar{F}^{-1}(c)}, \qquad \text{for all } w \in [c, 1] \text{ and } B \geq 0$$

Interestingly, this lower bound suggests that retailers selling low-margin products (those with large c) should receive less financial support from the supplier than retailers selling more profitable products.

Corollary 11.1 *Under the assumptions in Proposition 11.3, the retailer's initial payment to the supplier is equal to:*

$$\alpha_{\rm I}(w)\, w\, Q_{\rm I}(w, \alpha_{\rm I}(w)) = \begin{cases} B & \text{if } B \leq w\,\bar{F}^{-1}(w) \\ w\,\bar{F}^{-1}(w) & \text{otherwise.} \end{cases}$$

Proof: Follows directly from Proposition 11.3 and it is omitted.

According to Corollary 11.1, except when the retailer's budget is large ($B \geq w\,\bar{F}^{-1}(w)$) the supplier is willing to offer financial support (by setting $\alpha_{\rm I} < 1$) to induce the retailer to expend all his budget at time 0. Of course, by doing so the supplier is also increasing his credit risk exposure since $w\,Q_{\rm I} \geq B$ in these cases.

Let us now discuss the optimal choice of $w_{\rm I}(B)$ for the case $B > 0$. This condition is imposed to ensure that there exists a unique solution $\tilde{Q}(w, B)$ to $\bar{F}(\tilde{Q}) = w\,\bar{F}(w\,\tilde{Q} - B)$ for any $w \in [c, 1]$. The special case $B = 0$ is discussed in detail in Section 11.4.3.

Proposition 11.4 *Let* $B > 0$ *and suppose F satisfies the conditions in Assumption 11.1. Let* $\tilde{Q}(w, B)$ *be the unique solution to* $\bar{F}(\tilde{Q}) = w\,\bar{F}(w\,\tilde{Q} - B)$ *and define* $\dot{w} \in [c, 1]$ *such that* $\bar{F}(\tilde{Q}(\dot{w})) = c$. *Then, the optimal wholesale price* $w_{\rm I}(B)$ *is*

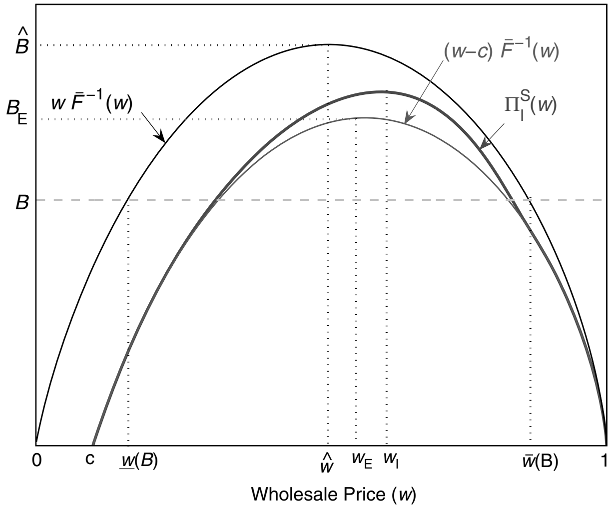


FIGURE 11.2 Supplier's expected payoff $\Pi_I^S(w)$ as function of the wholesale price w given a fixed retailer's budget B.

bounded below by $\dot{w}$ and solves:

$$w_I(B) = \arg\max_{\dot{w} \le w \le 1} \Pi_I^S(w, B) = \left\{ (w - c)\, \tilde{Q}(w, B) - \mathbb{E}\left[\left(D + B - w\, \tilde{Q}(w, B) \right)^{-} \right] \right\}$$

The retailer's optimal ordering level is $Q_I(B) = \tilde{Q}(w_I(B), B)$ and the optimal credit parameter is $\alpha_I(B) = \min\{1, B/(w_I(B)\, Q_I(B))\}$. The supplier's optimal expected payoff $\Pi_I^S(w_I(B), B)$ is monotonically decreasing in B.

Proof: See the Appendix at the end of the chapter.

In general, the optimal wholesale price $w_I(B)$ cannot be computed in closed-form, however, it can be easily found numerically. Nevertheless, Proposition 11.4 leads to a number of useful properties of the resulting equilibrium. For instance, it shows that the supplier prefers to do business with small budget retailers. The intuition behind this result is again driven by the retailer's limited liability that induces small budget retailers to take more risks by ordering more units compared to large budget retailers.

Figure 11.2 depicts the supplier's expected payoff $\Pi_I^S(w)$ as well as two auxiliary functions $w\,\bar{F}^{-1}(w)$ and $(w - c)\,\bar{F}^{-1}(w)$ that will be useful for the discussion of the following properties of an optimal solution (we refer the reader to Section 11.2.1 for the definitions of the notation used in the figure).

- Suppose $B \ge \hat{B}$ or $c \ge \bar{w}(B)$. It follows from Lemma 11.1 that $\tilde{Q}(w) = \bar{F}^{-1}(w)$ for all $w \in [c, 1]$. In this case, the retailer has a sufficiently large budget and the deposit contract coincides with the traditional wholesale contract with no budget constraint. That is, $\alpha_I(B) = 1$, $w_I(B) = w_E$ and $Q_I(B) = Q_E = \bar{F}^{-1}(w_E)$.

- Suppose $B < \hat{B}$ and $c < \bar{w}(B)$. In this case, we can get upper and lower bounds on w_I. First, after some straightforward manipulations we can rewrite the supplier's expected payoff as follows:

$$\Pi_I^S(w, B) = (w\,\tilde{Q} - B)\,\bar{F}(w\,\tilde{Q} - B) + \int_0^{(w\,\tilde{Q}-B)^+} x\mathrm{d}F(x) - c\,\tilde{Q} + B. \tag{11.7}$$

  Combining the following facts: (*i*) the function $x\,\bar{F}(x) + \int_0^x y\,f(y)\,\mathrm{d}y$ is increasing in x, (*ii*) $\tilde{Q}$ is a decreasing function of w, and (*iii*) $w\,\tilde{Q}$ is a unimodal function of w,[4] we conclude that $\Pi^S(w, B)$ is increasing in the range $[c, w_0(B)]$ where $w_0(B) = \text{argmax}\{w\,\tilde{Q}(w, B)\}$. It follows from the proof of Lemma 11.1 that $\tilde{Q}(w_0(B)) = \hat{Q} = \bar{F}^{-1}(\hat{w})$. As a result, $w_0(B)$ solves the equation $\hat{w} = w_0\,\bar{F}(w_0\,\bar{F}^{-1}(\hat{w}) - B)$ and so $w_0(B) \geq \hat{w}$. One can also show from the previous equation and the definition of $\underline{w}(B)$ that $w_0(B) \geq \underline{w}(B)$. We conclude that $w_I(B) \geq \max\{c, w_0(B)\} \geq \max\{c, \hat{w}, \underline{w}(B)\}$.
  On the other hand, a simple upper bound on the optimal wholesale price is given by $w_I \leq \max\{\bar{w}, w_E\} = w_T$, the wholesale price under a traditional wholesale contract with the budget constraint (see Proposition 11.1). It follows that the retailer is better off using if the supplier offers a credit contract instead of the traditional wholesale contract.
- Condition $\alpha_I = B/(w_I\,Q_I)$ together with Proposition 11.3 imply that at optimality $Q_I \leq Q_C$, where $Q_C = \bar{F}^{-1}(c)$ is the optimal centralized ordering quantity. Hence, the double marginalization inefficiency persists under a credit contract for all values of $B > 0$.

We conclude this section with the following result that provides a partial characterization of the notion of a "large budget," i.e., a budget $\bar{B}$ such that the credit contract equilibrium is invariant for $B \geq \bar{B}$.

Proposition 11.5 *Suppose the density function of D satisfies $f(0) = 0$. Then, $\bar{B} = B_E$ and $(w_I(B), Q_I(B) = (w_E, Q_E)$ for all $B \geq B_E$. In addition, the optimal wholesale price $w_I(B)$ is continuous in B. On the other hand, if $f(0) > 0$ then $B_E \leq \bar{B} \leq \hat{B}$.*

Proof: See the Appendix at the end of the chapter.

The following corollary follows directly from Corollary 11.1 and Proposition 11.5.

Corollary 11.2 *Suppose $f(0) = 0$. Then, for $B \leq B_E$ the retailer utilizes all his budget at time 0, that is, $\alpha_I(B)\,w_I(B)\,Q_I(B) = B$.*

[4] See Lemma 11.1 for a proof of (*ii*) and (*iii*).

11.4.3 SPECIAL CASE: $B = 0$

Let us discuss in more detail the case in which the retailer has no initial budget (i.e., $B = 0$). This is an important extreme case as it allows us to isolate the benefits of a credit contract in a situation in which the use of a traditional wholesale contract would lead to a nonoperative supply chain.

From Proposition 11.4, and the fact for $B = 0$ the ordering level $\tilde{Q}$ satisfies $\bar{F}(\tilde{Q}) = w\,\bar{F}(w\,\tilde{Q})$, it follows that the optimal wholesale price solves:

$$w_{\text{I}} = \underset{c \le w \le 1}{\operatorname{argmax}} \left\{ \tilde{Q}\,(\bar{F}(\tilde{Q}) - c) + \int_0^{w\,\tilde{Q}} D\,\mathrm{d}F(D) \right\}$$

Note that for $w = 1$, the equation $\bar{F}(\tilde{Q}) = w\,\bar{F}(w\,\tilde{Q})$ is satisfied for all $\tilde{Q} \ge 0$. This multiplicity of solutions follows from the fact that for $w = 1$ and $B = 0$ the retailer's payoff is identically 0 for any non-negative $\tilde{Q}$. We consider two alternative solutions to address this problem.

1. **Cooperative Retailer:** In this case, we assume that the retailer is indifferent among all possible values of $\tilde{Q} \ge 0$ when the supplier sets a wholesale price $w_{\text{I}} = 1$. It follows that the supplier can achieve the same outcome (and profit) as the centralized system. Indeed, it is not hard to see that it is in the supplier's best interest to set $w = 1$ and to induce the retailer to order a quantity $\tilde{Q}$ that maximizes the supplier's payoff:

$$\max_{\tilde{Q} \ge 0} \left\{ \tilde{Q}\,\bar{F}(\tilde{Q}) + \int_0^{\tilde{Q}} D\,\mathrm{d}F(D) - c\,\tilde{Q} \right\} = \max_{\tilde{Q} \ge 0} \left\{ \mathbb{E}[\min\{D, \tilde{Q}\}] - c\,\tilde{Q} \right\}$$

   This is exactly the payoff of a centralized system. The corresponding optimal credit contract satisfies $w_{\text{I}} = 1$, $\alpha_{\text{I}} = 0$, $Q_{\text{I}} = Q_{\text{C}} = \bar{F}^{-1}(c)$, $\Pi_{\text{I}}^{\text{R}} = 0$, and $\Pi_{\text{I}}^{\text{S}} = \Pi^{\text{C}} = \int_0^{\bar{F}^{-1}(c)} D\,\mathrm{d}F(D)$.

2. **Noncooperative Retailer:** In this case, rather than assuming that the retailer is indifferent and willing to order any quantity if $w = 1$, we assume that the retailer ordering quantity, $\tilde{Q}(w)$, is continuous on w so that $\tilde{Q}(1) = \lim_{w \uparrow 1} \tilde{Q}(w)$. It follows then that $\tilde{Q}(1) = \hat{Q}$ (recall that $\hat{Q}$ is the unique solution to $Q\,h(Q) = 1$).[5]

   The continuity of $\tilde{Q}(w)$ (together with its monotonicity, see Lemma 11.1) allows us to invert this function and write the supplier expected payoff as a function of $\tilde{Q}$ instead of w. For this, we must express w as a function of $\tilde{Q}$ solving the equation $\tilde{Q}\,\bar{F}(\tilde{Q}) = w\,\tilde{Q}\,\bar{F}(w\,\tilde{Q})$. By Assumption 11.1, the function $Q\,\bar{F}(Q)$ is unimodal and achieves its maximum

[5] An alternative way of modeling the behavior of a noncooperative retailer would be to assume that the retailer does not operate (i.e., selects $\tilde{Q} = 0$) if the supplier charges the wholesale price $w = 1$. However, from the discussion that follow, we can show that under this assumption the supplier optimization problem is ill-posed in the sense that he would like to select a wholesale price that is strictly less but as closed as possible to 1.

at $\hat{Q}$. This together with the fact that $w \leq 1$ implies that the function $w(\tilde{Q})$ is defined for $\tilde{Q} \geq \hat{Q}$. Hence, as a function of $\tilde{Q}$, the supplier maximizes his payoff solving:

$$\max_{\tilde{Q} \geq \hat{Q}} \Pi_I^S(\tilde{Q}) = \left\{ \tilde{Q}\,(\bar{F}(\tilde{Q}) - c) + \int_0^{w(\tilde{Q})\,\tilde{Q}} D\,\mathrm{d}F(D) \right\}$$

The unimodality of $Q\,\bar{F}(Q)$ together with the condition $\tilde{Q}\,\bar{F}(\tilde{Q}) = w(\tilde{Q})\,\tilde{Q}\,\bar{F}(w(\tilde{Q})\,\tilde{Q})$ imply that the function $w(\tilde{Q})\,\tilde{Q}$ is also unimodal on $\tilde{Q}$ attaining its maximum at $\hat{Q}$. As a result, Π_I^S is maximized at $\tilde{Q} = \hat{Q}$, which implies $w(\hat{Q}) = 1$. In summary, the outcome of the game in this case is given by $w_I = 1$, $\alpha_I = 0$, $Q_I = \hat{Q}$, $\Pi_I^R = 0$ and $\Pi_I^S = \hat{Q}\,[\bar{F}(\hat{Q}) - c] + \int_0^{\hat{Q}} D\,\mathrm{d}F(D)$.

If we compare the equilibrium outcomes of these two cases, we see that when $B = 0$ the supplier is always better off charging a full wholesale price ($w = 1$) independently of whether the retailer is cooperative or not and, as a result, the retailer ends up making no profit.

The main difference between these two cases is the ordering level and corresponding payoff of the supplier. On one hand, the cooperative retailer is willing to order, $Q_I = Q_C$, a quantity that gives the supplier the same payoff as the centralized system. In other words, the cooperative retailer is essentially transferring his retail business to the supplier at no cost. On the other hand, the non-cooperative retailer chooses a quantity $Q_I = \hat{Q}$, which is in general suboptimal from the supplier's point of view, $\Pi_I^S(\hat{Q}) \leq \Pi_I^S(Q_C)$. As for the ordering quantity, whether $Q_C \leq \hat{Q}$ or $Q_C \geq \hat{Q}$ depends on the production cost c and the demand distribution F. Indeed, $Q_C \leq \hat{Q}$ if and only if $\bar{F}^{-1}(c)\,h(\bar{F}^{-1}(c)) \leq 1$. This is an example of a situation in which decentralization increases the market output with the corresponding benefits for the end consumers.

11.5 Equilibrium with External Financing (Q_E, w_E)

In this section we consider an alternative mode of financing for the retailer. Instead of negotiating the term of a credit contract with the supplier, as in the previous section, we assume that the retailer gets a loan from a financial institution (e.g., a commercial bank) to fund his operation. The retailer uses this loan together with his initial budget to pay the supplier. To contrast the effects of this type of *external financing* with the one discussed in the previous section, we assume that the supplier offers a traditional wholesale contract that forces the retailer to pay the supplier in full at time 0 (i.e., $\alpha = 1$).

We represent the terms of the loan in the form of a triplet $(L, \mathcal{L}, r)$ where L is the amount that the retailer borrows at time 0, r is the nominal interest rate charged by the financial institution at time T and $\mathcal{L}$ are the earnings (possibly random) that the retailer has available at time T to repay the loan plus the interests. That is, the retailer is required to pay $\min\{\mathcal{L}, L(1+r)\}$ at the end of the selling season when demand uncertainty is resolved and revenues are collected. Of course, r should depend on L and $\mathcal{L}$ to capture the underlying market price of risk in the economy. To model this dependency we assume that the financial market is competitive in the following sense.

Definition 11.1 Competitive Financial Market: *We say that the financial market is competitive if the terms of a loan* $(L, \mathcal{L}, r)$ *satisfy*

$$L(1+r_f) = \mathbb{E}[\min\{\mathcal{L}, L(1+r)\}]$$

where r_f *is the risk-free rate.*

Intuitively, this condition says that on average (taking into account the risk of default) the financial market lends money at the risk-free rate r_f. Without loss of generality, in what follows we will normalize interest rates in this economy so that $r_f = 0$.

According to this definition, it should be clear that some loans $(L, \mathcal{L}, r)$ are infeasible in a competitive financial market. The following lemma provides a necessary and sufficient condition that ensures that a loan $(L, \mathcal{L}, r)$ can be effectively negotiated.

Lemma 11.2. *In a competitive financial market, a loan* $(L, \mathcal{L}, r)$ *is feasible if and only if* $\mathbb{E}[\mathcal{L}] \geq L$.

Proof: See the Appendix at the end of the chapter.

Let us now discuss the retailer's problem under this alternative source of (external) financing. After observing the wholesale price w offered by the supplier, the retailer must determine the quantity Q to order and the size L of the loan needed to fund this ordering level. For a given order quantity Q, and given the retailer's initial budget B, the retailer will always choose to minimize the size of the loan, that is, $L = (wQ - B)^+$. This is a consequence of two facts: (*i*) the interest rate increases with the size of the loan and (*ii*) our assumption that the retailer has no other investment opportunity besides his retail business (and possibly a cash account that pays the risk-free interest rate). As a result, we can write the retailer's optimization problem exclusively in terms of Q.

According to Lemma 11.2, the retailer has limited access to the credit market. Indeed, to get a loan $L = (wQ - B)^+$ the retailer's revenues at time T must satisfy $\mathbb{E}[\min\{D, Q\}] \geq (wQ - B)^+$. This condition defines an upper bound $\bar{Q}$ on the quantity that the retailer can order. For all $Q \leq \bar{Q}$, the interest rate $r(Q)$ that the retailer pays to get the loan $L = (wQ - B)^+$ is the unique solution to the equation:

$$\mathbb{E}\left[\min\{\min\{D, Q\} + B - wQ, (wQ - B)^+ r(Q)\}\right] = 0$$

Combining these conditions we get that the retailer's optimization problem is given by:

$$\Pi_E^R = \max_{0 \le Q \le \bar{Q}} \mathbb{E}\left[\left(\min\{D, Q\} + B - w\,Q - (w\,Q - B)^+ r(Q)\right)^+\right] - B \quad (11.8)$$

$$\text{subject to } \mathbb{E}\left[\min\{\min\{D, Q\} + B - w\,Q, (w\,Q - B)^+ r(Q)\}\right] = 0 \quad (11.9)$$

As before, Π^R denotes the retailer's expected payoff *net* of his initial budget B. We can solve the retailer's problem in a rather simple way. Indeed, if we use the identity $(x - y)^+ = x - \min\{x, y\}$ the retailer's objective can be rewritten as follows:

$$\begin{aligned}\Pi_E^R = \max_{0 \le Q \le \bar{Q}} & \mathbb{E}\left[\min\{D, Q\} + B - w\,Q\right] \\ & -\mathbb{E}\left[\min\{\min\{D, Q\} + B - w\,Q, (w\,Q - B)^+ r(Q)\}\right] - B\end{aligned}$$

And so by virtue of constraint (11.9) the retailer's optimization reduces to:

$$\Pi_E^R = \max_{0 \le Q \le \bar{Q}} \mathbb{E}\left[\min\{D, Q\} - w\,Q\right]$$

$$\text{subject to } \mathbb{E}\left[\min\{\min\{D, Q\} + B - w\,Q, (w\,Q - B)^+ r(Q)\}\right] = 0$$

Note that the objective function corresponds to the standard payoff of the retailer under a wholesale contract (see the discussion at the end of Section 11.2). In particular, this objective is independent of the interest rate $r(Q)$ and the retailer's initial budget B. The optimal ordering level is $\bar{F}^{-1}(w)$ that trivially satisfies the feasibility constraint $\bar{F}^{-1}(w) \le \bar{Q}$.

It is worth noticing that the retailer optimal solution is independent of B and coincides with the one that is obtained in the traditional wholesale contract with no budget constraint. In order words, access to a competitive financial market has enabled the retailer to completely separate the operation of his retail business (in particular his procurement decisions) from its financing. The following proposition summarizes this solution.

Proposition 11.6 *Suppose the retailer has an initial budget B and has access to a competitive financial market. Then, under a wholesale contract with wholesale price w, the retailer's optimal strategy is to order a quantity $\bar{F}^{-1}(w)$ and to get a loan (if necessary) for an amount equal to $L = (w\,\bar{F}^{-1}(w) - B)^+$. The interest rate, r, that the retailer pays on this loan is the unique solution to*

$$\mathbb{E}\left[\min\{\min\{D, \bar{F}^{-1}(w)\} + B - w\,\bar{F}^{-1}(w), (w\,\bar{F}^{-1}(w) - B)^+ r\}\right] = 0.$$

Given the retailer's best strategy $\bar{F}^{-1}(w)$, the supplier problem reduces to

$$\Pi_E^S = \max_{c \le w \le 1} \left\{(w - c)\,\bar{F}^{-1}(w)\right\}$$

with optimal solution w_E. As a result, the optimal ordering level is $Q_E = \bar{F}^{-1}(w_E)$.

Based on our discussion about the traditional wholesale contract at the end of Section 11.2, it follows that $w_E = \lim_{B\to\infty} w_T(B)$ and $Q_E = \lim_{B\to\infty} Q_T(B)$. In other words, the existence of a competitive financial market allows the entire supply chain to operate in exactly the same way as it would operate if the retailer had no budget constraint.

11.6 Computational Experiments

In this section we conduct a set of numerical computations to compare the equilibrium under *internal* financing of Section 11.4 and the equilibrium under *external* financing of the previous section. We do not attempt to have an exhaustive set of experiments that describe all possible outcomes of the game. Instead, we have chosen a subset of instances that highlight some of the most important features of the equilibrium and the main differences between the alternative forms of financing.

To perform these experiments we use a (λ, k)-Weibull distribution[6] to model demand uncertainty. The reasons for this choice are (*i*) the Weibull satisfies Assumption 11.1, (*ii*) it simplifies the computation of the equilibria of the game, and (*iii*), as a two-parameter distribution, it offers enough flexibility to fit at least the first two moments of the demand. For the case $B = 0$, we have chosen the option of a *cooperative retailer* (see Section 11.4.3 for details).

Our first set of experiments compares the market equilibrium for different types of contracts as a function of the retailer's budget B. Figure 11.3 plots the wholesale price (left panel) and ordering level (right panel) for the credit contract $(w_I(B), Q_I(B))$, for the case of external financing $(w_E(B), Q_E(B))$, and for the traditional wholesale contract $(w_T(B), Q_T(B))$. The right panel also includes the centralized production quantity Q_C.

It follows from Figure 11.3 that all three modes of operations (credit contract, external financing, or traditional wholesale contract) are equivalent if the retailer's budget is sufficiently large, in this case $B \geq B_E$.[7] It is also worth noticing that under a traditional wholesale contract (i.e., when the retailer does not receive any type of financial support), the supplier charges a higher wholesale price and the retailer orders less compared to the cases with internal or external financing. Therefore, the supply chain (as a whole) is worst off if the retailer is unable to get financial support (either from the supplier or a third party financial institution).

If we compare the outcome under internal versus external financing, we see that under internal financing the retailer orders more units than with external financing. As a result, final consumers are better off under a credit contract. In terms of the wholesale price, there is no absolute comparison between the two types of financing. If B is relatively small, the credit contract leads to higher wholesale prices but if B gets large then external financing leads to higher wholesale prices.

[6] The tail distribution of a (λ, k)-Weibull is given by $\bar{F}(x) = \exp\left(-(x/\lambda)^k\right)$. Its mean and variance are given by $\mu = \lambda\,\Gamma\left(1+\frac{1}{k}\right)$ and $\sigma^2 = \lambda^2\left[\Gamma\left(1+\frac{2}{k}\right) - \Gamma^2\left(1+\frac{1}{k}\right)\right]$, respectively.

[7] This follows from Propositions 11.1 and 11.5 and the fact that $f(0) = 0$ in this example.

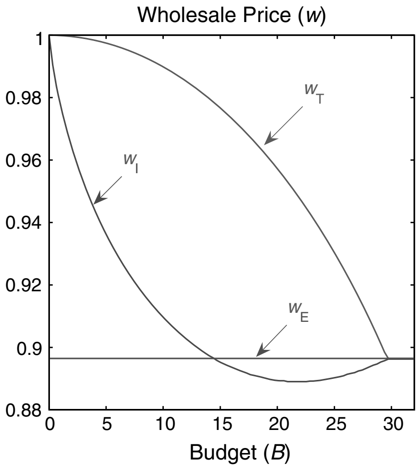


FIGURE 11.3 **Equilibrium wholesale price (left panel) and ordering level (right panel) as a function of the retailer's initial budget (B) under three alternative contracts: credit contract (w_{I}, Q_{I}), external financing (w_{E}, Q_{E}) and traditional wholesale contract (w_{T}, Q_{T}). Data: $c = 0.7$ and $\bar{F}(x) = \exp\big(-(0.01\,x)^2\big)$.**

Hence, from the point of view of the retailer neither mode of financing dominates the other, that is, his preferences depend on his initial budget.

Figure 11.4 shows the retailer (left panel) and supplier (right panel) expected payoff for the three contracts. It follows that the supplier is always better off under the credit contract independently of the retailer's initial budget. One of the remarkable features of the credit contract is that the retailer's expected payoff $\Pi_{\mathrm{I}}^{\mathrm{R}}(B)$ is non-monotonic on his initial budget B. There is a finite budget, B_m in Figure 11.4, at which the retailer's payoff is maximized (for all three types of

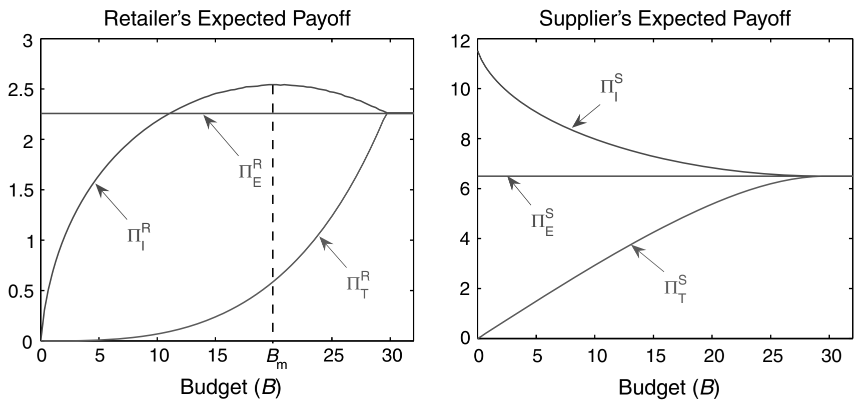


FIGURE 11.4 **Retailer's payoff (left panel) and supplier's payoff (right panel) as a function of the retailer's initial budget. Data: $c = 0.7$ and $\bar{F}(x) = \exp\big(-(0.01\,x)^2\big)$.**

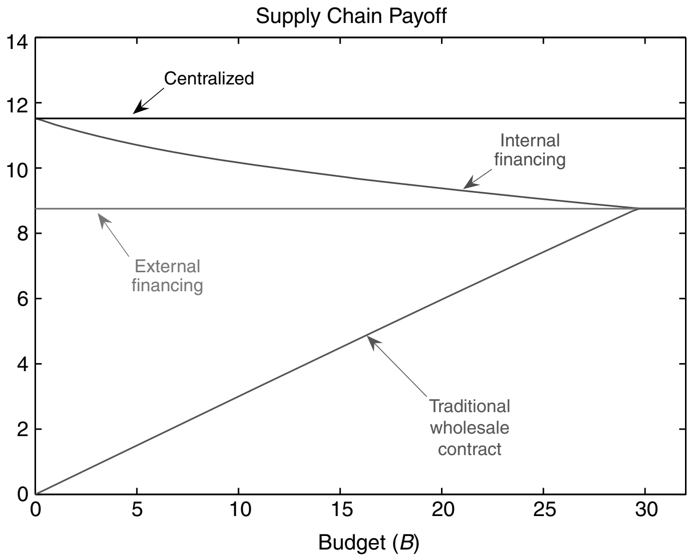


FIGURE 11.5 Supply chain payoff under different modes of operations. Data: $c = 0.7$ **and** $\bar{F}(x) = \exp\big(-(0.01\,x)^2\big)$.

contracts considered in our discussion). This non-monotonic feature of $\Pi_{\mathrm{I}}^{\mathrm{R}}(B)$ suggests that implementing a credit contract can be difficult in practice since in those cases in which $B \geq B_m$ the retailer has incentives to underreport his true budget at the time the credit contract is negotiated. In Section 11.7, we extend our model to the case in which there is asymmetric information between the retailer and supplier about the true value of B. Interestingly, since the supplier's payoff under a credit contract, $\Pi_{\mathrm{I}}^{\mathrm{S}}(B)$, is monotonically decreasing in B, then for $B \geq B_m$ both the retailer and the supplier would be better off if the retailer had a smaller initial budget. It also follows from Figure 11.4 that both agents are worst off if they operate using a traditional wholesale contract. Hence, it is in the retailer and supplier best interests to get some type of financing for the retailer.

As Figure 11.5 reveals, the overall payoff of the entire supply chain is monotonically decreasing in B under a credit contract. Despite this fact, the credit contract is the best mode of operations for the entire supply chain independent of the value of B. On the contrary, the traditional wholesale contract produces the worst possible outcome for the aggregate system (including consumers surplus).

Our next computational experiment highlights a curiosity of the equilibrium under a credit contract. The left panel in Figure 11.6 plots the wholesale price for the three type of contracts as a function of the retailer's budget. As we can see, in this example the wholesale price for the credit contract $w_{\mathrm{I}}(B)$ has a discontinuity at $B_d \approx 36$. In particular, $w_{\mathrm{I}}(B)$ is monotonically decreasing for $B \in [0, B_d)$ jumps upward at B_d and remains constant for $B \geq B_d$. To understand the nature of this discontinuity, the right panel in Figure 11.6 plots the supplier's payoff $\Pi_{\mathrm{I}}^{\mathrm{S}}(w, B)$ as a function of w for four values of B (with $B_a < B_b < B_c < B_d$). For $B = B_d$,

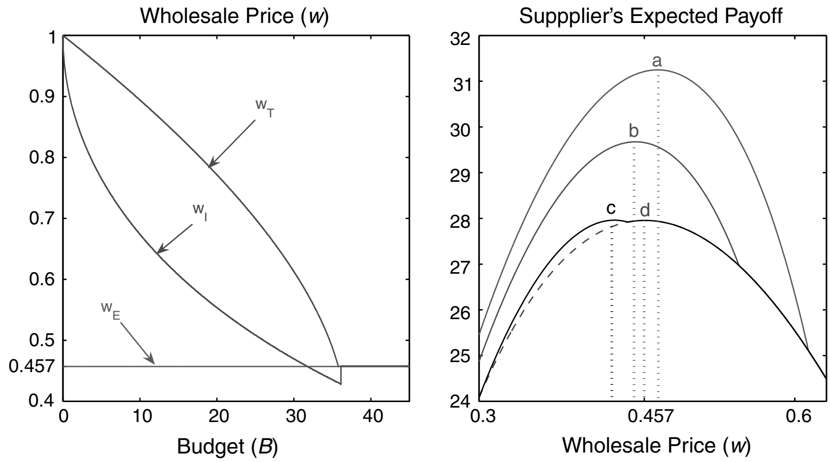


FIGURE 11.6 **Wholesale price (left panel) and Supplier's Payoff (right panel). Data:** $c = 0.1$ **and** $\bar{F}(x) = \exp\big(-0.01\,x\big)$.

the budget is sufficiently large and it is not a constraint, as a result $\Pi_{\text{I}}^{\text{S}}(w, B_d) = (w - c)\,\bar{F}^{-1}(w)$ (this curve is represented by the dashed line in the figure). As we can see, the optimal wholesale price decreases from a to b and from b to c and then jumps from c to d. This discontinuity on the optimal wholesale price implies that the retailer's expected payoff (as well as the payoff of the entire supply chain) has a discontinuity at B_d, specifically a downward jump. It is worth mentioning that this discontinuity does not contradict the result in Proposition 11.5 since in the example in Figure 11.6 demand has an exponential distribution and so $f(0) > 0$. This is in contrast to the example in Figure 11.3 for which demand has a Weibull distribution (with $f(0) = 0$) and the wholesale price $w_{\text{I}}(B)$ is continuous.

We conclude this section measuring the impact of demand variability on the performance of the three contracts. Table 11.2 shows the retailer and supplier payoffs for the external financing contract and traditional wholesale contract relative to their payoffs under a credit contract as a function of the retailer's budget (B) and the coefficient of variation (StDev/Mean) of the market demand. To isolate the effect of demand variability, we keep the mean demand constant and vary the standard deviation.

From the point of view of the retailer, the option of external financing is generally the best for all levels of demand variability. The credit contract comes next and the traditional wholesale contract is the worst. On the other hand, for the supplier the credit contract generally is the the best option except when the budget is small and the variability is medium to high ($B \leq 5$ and CV = 200%) in which case the external financing gives the supplier a higher payoff. In other words, suppliers serving small retailers facing high demand variability are better off if a bank takes the risk of providing financial services to the retailer. Similarly to the retailer, the traditional wholesale contract is the worst option for the supplier. We can also see from Table 11.2 that all three contracts produce the same outcome

TABLE 11.2 Retailer and Supplier Payoffs for the External Financing and Traditional Wholesale Contracts Relative to Their Payoffs under a Credit Contract as a Function of the Retailer's budget (*B*) and the Coefficient of Variation (StDev/Mean) of the Market Demand*

Retailer and Supplier Normalized Payoffs					
	Coefficient of Variation (%)				
B	10%	50%	100%	200%	500%
	External Financing Contract (Π_E^R/Π_I^R, Π_E^S/Π_I^S)				
1	(178.5 , 93.7)	(294.3 , 65.4)	(735.0 , 49.2)	(2170.2 , 342.1)	(100.0 , 100.0)
5	(112.9 , 98.0)	(146.0 , 76.0)	(233.5 , 55.2)	(149.8 , 104.7)	(100.0 , 100.0)
10	(102.9 , 99.4)	(113.8 , 84.2)	(136.8 , 63.2)	(100.1 , 100.0)	(100.1 , 100.0)
25	(100.1 , 100.0)	(94.7 , 98.0)	(100.0 , 100.0)	(100.1 , 100.0)	(100.1 , 100.0)
50	(99.9 , 100.0)	(99.9 , 100.0)	(100.0 , 100.0)	(100.1 , 100.0)	(100.3 , 100.0)
100	(99.9 , 100.0)	(99.9 , 100.0)	(100.0 , 100.0)	(100.1 , 100.0)	(100.5 , 100.0)
250	(99.9 , 100.0)	(99.9 , 100.0)	(100.0 , 100.0)	(100.1 , 100.0)	(101.3 , 100.0)
	Traditional Wholesale Contract (Π_T^R/Π_I^R, Π_T^S/Π_I^S)				
1	(0.0 , 1.6)	(0.0 , 2.4)	(1.4 , 4.2)	(100.3 , 100.0)	(100.0 , 100.0)
5	(0.0 , 8.3)	(0.2 , 14.2)	(12.4 , 22.2)	(100.1 , 100.0)	(100.0 , 100.0)
10	(0.0 , 16.8)	(1.7 , 31.1)	(32.2 , 46.1)	(100.1 , 100.0)	(100.1 , 100.0)
25	(0.0 , 42.4)	(28.3 , 83.9)	(100.0 , 100.0)	(100.1 , 100.0)	(100.1 , 100.0)
50	(4.1 , 84.6)	(99.9 , 100.0)	(100.0 , 100.0)	(100.1 , 100.0)	(100.3 , 100.0)
100	(99.9 , 100.0)	(99.9 , 100.0)	(100.0 , 100.0)	(100.1 , 100.0)	(100.5 , 100.0)
250	(99.9 , 100.0)	(99.9 , 100.0)	(100.0 , 100.0)	(100.1 , 100.0)	(101.3 , 100.0)

*Demand has a Weibull distribution with fixed mean $\mu = 80$ and the manufacturing cost is $c = 0.5$.

when demand variability is high (CV = 500%) independent of the budget. The reason for this is that (for a Weibull distribution) the retailer's optimal ordering quantity converges to 0 as the demand variability increases.

In Table 11.3 we compare the efficiency of the three contracts from the point of view of the entire supply chain. For each type of contract we compute the percentage competition penalty as a function of the retailer's budget *B* and the coefficient of variation of the market demand:

$$\mathcal{P} := 1 - \frac{\Pi^R + \Pi^S}{\Pi^C} \times 100\%$$

As we can see from Table 11.2, the credit contract is generally the most efficient of the three contracts while the traditional wholesale contract is the less efficient. Again, for the case when the budget is small and the variability is medium to high ($B \leq 5$ and CV = 200%) the option of external financing is the most efficient. We also note that the efficiency of the credit contract decreases

TABLE 11.3 Competition Penalty as a Function of the Retailer's Budget (B) and the Coefficient of Variation (StDev/Mean) of the Market Demand*

PERCENTAGE COMPETITION PENALTY ($\mathcal{P}$)					
	Coefficient of Variation (%)				
B	10%	50%	100%	200%	500%
			Credit Contract		
1	11.2	8.0	0.5	86.2	26.7
5	12.9	13.2	3.3	39.0	26.7
10	13.5	16.6	6.2	26.9	26.7
25	13.8	22.4	26.6	26.9	26.7
50	13.8	24.6	26.6	26.9	26.8
100	13.8	24.6	26.6	26.9	26.9
250	13.8	24.6	26.6	26.9	27.1
			External Financing		
For any B	13.8	24.7	26.6	26.9	26.7
			Traditional Wholesale Contract		
1	98.6	97.9	95.9	86.2	26.7
5	93.2	89.6	79.6	39.0	26.7
10	86.4	79.1	59.4	26.9	26.7
25	66.1	46.4	26.6	26.9	26.7
50	32.1	24.7	26.6	26.9	26.7
100	13.8	24.7	26.6	26.9	26.7
250	13.8	24.7	26.6	26.9	26.7

*Demand has a Weibull distribution with fixed mean $\mu = 80$ and the manufacturing cost is $c = 0.5$.

with the retailer's budget B and the opposite is true for the traditional wholesale contract. Interestingly, for small values of B, the efficiency of the credit contract is U-shape as a function of the level of variability of the demand. In these cases, the competition penalty is maximized at intermediate values of the coefficient of variation. However, as B increases the efficiency of the credit contract becomes monotonically decreasing in the level of demand uncertainty.

11.7 Concluding Remarks and Extensions

In this chapter we investigate how financial constraints affect the operations of a two-echelon supply chain and discuss two alternative modes of financing that could be used to limit the effects of these constraints. Our model focuses on the case in which the retailer has a limited amount of working capital (his budget B)

to procure products from the supplier (alternatively, we could have considered the reversed situation in which the supplier is the small player; a scenario that could be more appropriate in certain industries).

We set our mathematical formulation within the framework of the newsvendor model, that is, we consider a single-period model in which the retailer procures products from the supplier to satisfy a future stochastic demand. We also restrict attention to simple (yet popular) linear transfer payments (or wholesale contracts) between the agents. In these contracts, the supplier first specifies a per unit wholesale price w and the retailer then selects the order size q. Typically, in the absence of a budget constraint, the retailer pays the full amount wq at the time the contract is signed. In our model, this traditional wholesale contract needs to be modified to include the additional budget constraint $wq \leq B$, which could seriously affect the profitability of the entire supply chain if B is small. To overcome this limitation, we discuss two possible types of financing for the retailer. First, we consider the case of *internal financing* in which the supplier charges only a fraction $\alpha \in [0, 1]$ of the wholesale price at the time the contract is signed. The remaining fraction $(1 - \alpha)$ is paid after market demand is realized. With this contract, the effective budget constraint at the moment the order is placed is $\alpha wq \leq B$ which, depending on the value of α, allows the retailer to increasing his order size. From the supplier standpoint, this *credit contract* increases the overall production of the supply chain helping his own business. On the downside, the supplier faces a new source of risk under a credit contract when the retailer defaults on his obligation to paid the remaining portion $(1 - \alpha)wq$ after demand is realized. Nevertheless, it is obvious that the supplier is always better off (on average) using the credit contract instead of a traditional wholesale contract since he can always set $\alpha = 1$.

The second form of financing that we consider is the retailer applying for a commercial loan from a financial institution (e.g., a bank). For the sake of simplicity, we assume that the interests charged on this loan are set based on a competitive financial market assumption (see Definition 11.1 for details). In this *external financing* case, the supplier offers a traditional wholesale contract requiring full payment at the time the contract is signed. In Section 11.5 we study this contract and show that its equilibrium is equivalent to the traditional wholesale contract in which the retailer has an unlimited budget. This conclusion is a reminiscent of the well-known Modigliani and Miller's irrelevance principle. An interesting extension regarding this part of our model would be to consider a less competitive financial market in which the interest rate are set using a different equilibrium concept (e.g., the CAPM model).

Our analysis and results for the credit contract are somehow less palpable. First of all, we show that for any given value of w and α the retailer's optimal ordering quantity is decreasing in B. That is, small retailer's are willing to order more (i.e., take more risk) since they have less stakes at risk in case of default. The implication of this behavior is that the optimal ordering quantity and the supplier's optimal expected payoff are also decreasing in the retailer's budget B. In other words, the supplier prefers to do business with a small retailer willing to take more risk than with a larger retailer (with more collateral) that is more cautious in his procurement decisions. In the extreme case in which $B = 0$ we show that, depending on the retailer's preferences, it is possible for the supplier to achieve the same expected payoff as in the centralized solution. Neither the traditional wholesale contract nor the external financing option are able to achieve this first best (actually the traditional wholesale contract leads to a nonoperative supply chain if $B = 0$). Another interesting feature of the credit contract is that the optimal wholesale price, $w_{\text{I}}(B)$, is a non-monotonic function of the retailer initial budget. Starting at $B = 0$, the wholesale price first decreases as B increases reaching a minimum at some value w_m when $B = B_m$ and then increasing (either smoothly or with an upward jump) to a value w_{E}. This value w_{E} is equal to the optimal wholesale price under external financing (or equivalently, under a traditional wholesale contract if retailer has sufficiently large budget).

Probably, one of the main assumptions that we make in our model is that the retailer's demand information and budget are public knowledge. Under this complete information assumption, the supplier can offer a credit contract ($\alpha(B)$, $w(B)$) specially designed and targeted to the retailer with initial budget B that maximizes the supplier's expected payoff. In practice, however, we do expect some degree of asymmetric information that limits the supplier's ability to taylor the credit contract using the retailer's initial wealth. In this setting, the supplier must design a contract that takes into account the *adverse selection* problem (e.g., Stiglitz and Weiss (1981)) under which the retailer (potentially) misrepresents his initial budget to improve the terms of the credit contract. We conclude this section measuring the impact of this behavior. However, we do not attempt a rigorous treatment of this adverse selection problem here as it goes beyond the scope of this chapter. We do show, however, a simple extension of the credit contract that can take care of this problem in certain cases.

To understand the nature of this adverse selection problem and get more insights about its implications on the actions of the retailer, let us consider the case in which the supplier offers naively the optimal credit contract menu $\{(\alpha_{\text{I}}(B), w_{\text{I}}(B)); B \geq 0\}$ that we derived in Proposition 11.4 under full information. The sequence of events is as follows: First, the supplier offers the menu of credit contracts $\{(\alpha_{\text{I}}(B), w_{\text{I}}(B)); B \geq 0\}$. The retailer then reports his initial budget $\tilde{B}$ and the specific contract $(\alpha_{\text{I}}(\tilde{B}), w_{\text{I}}(\tilde{B}))$ is selected from the menu. Finally, the retailer orders a quantity Q and pays the supplier the amount $\alpha_{\text{I}}(\tilde{B})\, w_{\text{I}}(\tilde{B})\, Q$ at time $t = 0$. The remaining portion $(1 - \alpha_{\text{I}}(\tilde{B}))\, w_{\text{I}}(\tilde{B})\, Q$ is paid in arrears at time T after market demand is realized. Because of the information asymmetry, the supplier has no concrete mechanism to verify that the retailer is effectively reporting his true initial wealth. Furthermore, to be consistent with our previous discussion, we assume that the retailer, after reporting his budget $\tilde{B}$, is free to choose any Q that satisfies the budget constraint. Naturally, at this point, the supplier could realize that the retailer is misreporting his budget if $Q \neq Q_{\text{I}}(\tilde{B})$, the optimal ordering level under full information. To avoid this problem, the supplier can modify the terms of the contract an offer a menu of *enlarged* credit contracts that also specifies the ordering quantity, that is, a menu $\{(\alpha_{\text{I}}(B), w_{\text{I}}(B), Q_{\text{I}}(B)); B \geq 0\}$. We will discuss this variation of the credit contract below. We also assume that if the retailer declares bankruptcy at time T then the supplier is able to audit the retailer's assets so that B becomes public knowledge at this point.

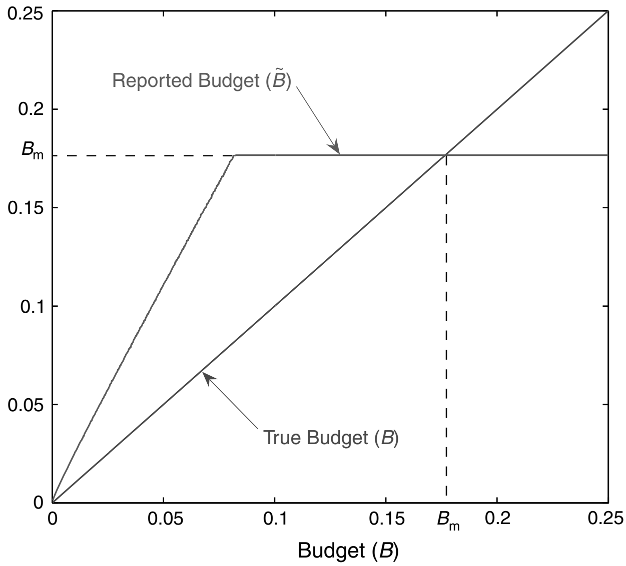


FIGURE 11.7 Retailer's true and reported budget. Demand follows a Uniform distribution in $[0, 1]$ **and** $c = 0.6$

Figure 11.7 plots the retailer's true budget (B) and reported budget ($\tilde{B}$) as a function of B for the case in which demand is uniformly distributed in $[0, 1]$. As we can see from the figure, when the retailer's true budget is below a threshold B_m the retailer has incentives to report a larger budget while when his budget exceeds B_m he has incentives to report a smaller budget. The threshold B_m is the one at which the wholesale price offered by the retail is minimized, that is, $B_m = \text{argmin}\{w_{\text{I}}(B)\,;\, B \geq 0\}$. Note that for small values of B the retailer reports a budget $\tilde{B}$ that is strictly less than B_m. The reason is that the credit parameter $\alpha_{\text{I}}(B)$ is in B. Hence, even if reporting B_m minimizes the wholesale price the corresponding budget constraint $\alpha_{\text{I}}(B_m)\, w_{\text{I}}(B_m)\, Q \leq B$ imposes a rather restrictive upper bound on the feasible ordering quantity Q.

The effect on the agents' payoffs of this misreporting can be significant. Figure 11.8 depicts the retailer (left panel) and supplier (right panel) expected payoffs as function of the retailer's true budget (B) for the cases in which the retailer reports his true budget (full information) and the case in which he misreports his budget (incomplete information). As expected, the retailer can take advantage of this asymmetry of information to increases his payoff. For the supplier, this misreporting can lead to a sharp decrease in his payoff, especially when B is small. It is interesting to note that from the point of view of the entire supply chain, the system is better off under incomplete information for $B \geq B_m$ while the system expected payoff is larger under full information for $B \leq B_m$.

As we mentioned before, the supplier can try to solve this adverse selection problem by including the quantity $Q_{\text{I}}(B)$ in the contract offering the menu $\{(w_{\text{I}}(B), \alpha_{\text{I}}(N), Q_{\text{I}}(B))\,;\, B \geq 0\}$. Based on our previous discussion, it is not hard

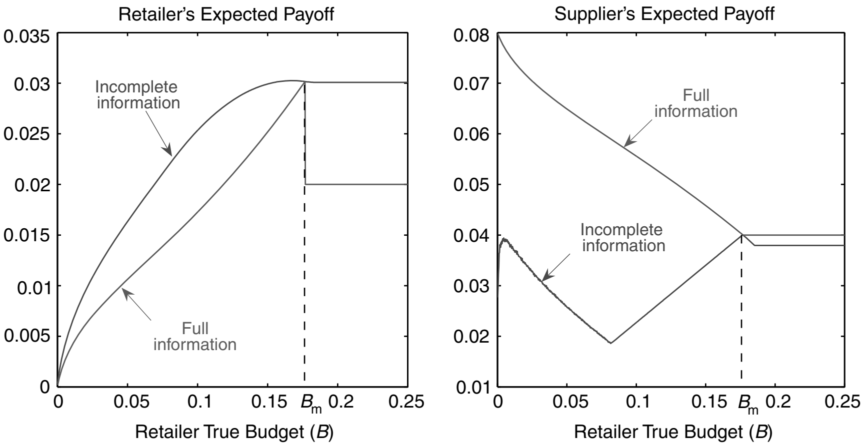


FIGURE 11.8 **Retailer (left panel) and supplier (right panel) expected payoffs as function of the retailer's true budget for the cases in which the retailer reports his true budget (full information) and the case in which he misreports his budget (incomplete information). Demand follows a Uniform distribution in** $[0, 1]$ **and** $c = 0.6$

to see that this modified contract takes care of the adverse selection problem only if $B \le B_m$. In this range, the retailer will truthfully declare his real budget. This follows from the fact that $\alpha_{\text{I}}(B)\, w_{\text{I}}(B)\, Q_{\text{I}}(B) = B$ in this range and so the retailer is unable to report a higher budget. However, the retailer will still underreport his budget if $B > B_m$ to get a lower wholesale price. To solve this problem, the supplier would have to offer some sort of compensation (information rents) to the retailer if he wants him to report his true budget. This compensation could come in the form of a transfer payment after demand is realized and payoffs are collected. An interesting extension to our model would be to formalize the supplier's mechanism design problem that could be used to identify such an optimal compensation scheme.

Appendix: Main Proofs

Proof of Lemma 11.1: We will prove that there is a unique solution to the first-order optimality condition:

$$\pi^{\text{R}}_{\text{Q}}(Q) \triangleq \frac{\mathrm{d}\pi^{\text{R}}(Q)}{\mathrm{d}Q} = \bar{F}(Q) - w\,\bar{F}(w\,Q - B) = 0$$

Note that $\pi^{\text{R}}(Q) \triangleq \mathbb{E}[(\min\{Q, D\} + B - w\,Q)^+ - B]$. Indeed, based on the inequality below it follows that a sufficient condition for the existence of $\tilde{Q}$ is that the equation $\bar{F}(Q) = w\,\bar{F}(w\,Q)$ admits a non-negative solution.

$$w\,\bar{F}(w\,Q - B) \ge w\,\bar{F}(w\,Q), \quad \text{for all } Q \ge 0$$

Suppose, by contradiction, that $\bar{F}(Q) > w\,\bar{F}(w\,Q)$, for all $Q \geq 0$. Integrating over $Q \in [0, \infty)$, we get:

$$\mathbb{E}[D] = \int_0^\infty \bar{F}(Q)\,\mathrm{d}Q > w \int_0^\infty \bar{F}(w\,Q)\,\mathrm{d}Q = \mathbb{E}[D]$$

This contradictions prove the existence of $\tilde{Q}$. To prove the uniqueness of $\tilde{Q}$ note that:

$$\begin{aligned}\frac{\mathrm{d}^2\pi^{\mathrm{R}}}{\mathrm{d}Q^2}\bigg|_{\tilde{Q}} &= -f(\tilde{Q}) + w^2 f(w\,\tilde{Q} - B)\\ &= -\bar{F}(\tilde{Q})\left[h(\tilde{Q}) - \frac{w^2\,\bar{F}(w\,\tilde{Q} - B)}{\bar{F}(\tilde{Q})}\,h(w\,\tilde{Q} - B)\right]\\ &= -\bar{F}(\tilde{Q})\left[h(\tilde{Q}) - w\,h(w\,\tilde{Q} - B)\right] < 0\end{aligned}$$

The last equality uses the fact that $\tilde{Q}$ satisfies $\pi^{\mathrm{R}}_{\mathrm{Q}}(\tilde{Q}) = 0$ and the inequality uses the fact that $w < 1$ and the hazard function $h(D)$ is strictly decreasing in D. As a result every local extrema $\tilde{Q}$ is a local maxima. By the continuity of π^{R} we conclude that there exists a unique $\tilde{Q}$ that solves the first-order optimality condition for π^{R}.

Let us now prove that $\tilde{Q}$ is decreasing in w. Recall the definitions of $\hat{w}$, $\bar{w}(B)$ and $\underline{w}(B)$ in Section 11.2.1. The following cases are possible:

- $c \geq \bar{w}$. In this case $\tilde{Q}(w) = \bar{F}^{-1}(w)$ for all $w \in [c, 1]$, which is decreasing in w.
- $c < \bar{w}$. For all $w \in [c, \underline{w}] \cup [\bar{w}, 1]$ we have that $\tilde{Q}(w) = \bar{F}^{-1}(w)$, again a decreasing function of w. So let us focus on the case $w \in (\underline{w}, \bar{w})$. In this case, $w\,\tilde{Q}(w) > B$ and we can use the identity $\bar{F}(\tilde{Q}(w)) = w\,\bar{F}(w\,\tilde{Q}(w) - B)$ to compute the derivative:

$$\begin{aligned}\frac{\mathrm{d}\tilde{Q}(w)}{\mathrm{d}w} &= \frac{\bar{F}(w\,\tilde{Q} - B) - w\,\tilde{Q}f(w\,\tilde{Q} - B)}{w^2 f(w\,\tilde{Q} - B) - f(\tilde{Q})}\\ &= \frac{1 - w\,\tilde{Q}\,h(w\,\tilde{Q} - B)}{w^2\,h(w\,\tilde{Q} - B) - w\,h(\tilde{Q})}, \quad w \in (\underline{w}, \bar{w})\end{aligned}$$

Note that the second equality uses the fact that $w\,\bar{F}(w\,\tilde{Q}(w) - B) = \bar{F}(\tilde{Q}(w))$. By Assumption 11.1 the hazard function $h(Q)$ is increasing, this together with the fact $w < 1$ imply that the denominator is negative. Hence, we would like to show that the nominator is non-negative. For this, we will prove the sufficient condition $1 - w\,\tilde{Q}\,h(w\,\tilde{Q}) \geq 0$. Let us define $w^0 = \operatorname{argmax}\{w\,\tilde{Q}(w) : w \in [0, 1]\}$, which satisfies the first-order optimality condition:

$$\frac{\mathrm{d}w\,\tilde{Q}(w)}{\mathrm{d}w}\bigg|_{w=w^0} = 0 \quad \text{which is equilvalent to} \quad \tilde{Q}(w^0)\,h(\tilde{Q}(w^0)) = 1$$

As a result, (by the monotonicity of $h(Q)$) we get that (as required):

$$1 - w\,\tilde{Q}(w)\,h(w\,\tilde{Q}(w)) \geq 1 - w^0\,\tilde{Q}(w^0)\,h(w^0\,\tilde{Q}(w^0)) \\ \geq 1 - \tilde{Q}(w^0)\,h(\tilde{Q}(w^0)) = 0, \quad \text{for all } w \in [\underline{w}, \bar{w}]$$

The monotonicity of $\tilde{Q}$ on B follows from noticing that $w\,\bar{F}\left(\frac{w\,Q-B}{p}\right)$ increases with B and remains constant at w for $B \geq w\,Q$.

Finally, the unimodality of $w\,\tilde{Q}(w, B)$ follows from the fact that $\tilde{Q}(w, B)$ solves the equation $\bar{F}(\tilde{Q}) = w\,\bar{F}(w\,\tilde{Q} - B)$. Taking derivative with respect to w (and after some straightforward manipulations) we get that:

$$\frac{\mathrm{d}(w\tilde{Q})}{\mathrm{d}w} = \frac{\tilde{Q}\,h(\tilde{Q}) - 1}{h(\tilde{Q}) - w\,h(w\,\tilde{Q} - B)}$$

But $\tilde{Q}$ is decreasing in w and by Assumption 11.1 $h(\tilde{Q})$ is increasing $\tilde{Q}$. It follows that $w\tilde{Q}$ is unimodal as a function of w and attains its maximum at a w_0 such that $\tilde{Q}(w_0, B)\,h(\tilde{Q}(w_0, B)) = 1$.

Proof of Proposition 11.2: Let us define $\pi^{\mathrm{R}}(Q) = \mathbb{E}[(\min\{Q, D\} + B - w\,Q)^+ - B]$ and note that $\Pi^{\mathrm{R}} = \max_Q \pi^{\mathrm{R}}(Q)$. We will show that $\pi^{\mathrm{R}}(Q)$ is unimodal in Q and that it admits the maximizer Q^* described in the statement of the proposition. Indeed:

$$\pi^{\mathrm{R}}_{\mathrm{Q}}(Q) \triangleq \frac{\mathrm{d}\pi^{\mathrm{R}}(Q)}{\mathrm{d}Q} = \begin{cases} \bar{F}(Q) - w & \text{if } w\,Q \leq B \\ \bar{F}(Q) - w\,\bar{F}(w\,Q - B) & \text{if } w\,Q \geq B \end{cases}$$

In the region $w\,Q \leq B$, the derivative $\pi^{\mathrm{R}}_{\mathrm{Q}}(Q)$ is decreasing in Q and so $\pi^{\mathrm{R}}(Q)$ is concave in this region. In the region $w\,Q \geq B$, it follows from the proof of Lemma 11.1 that $\pi^{\mathrm{R}}(Q)$ is unimodal in this region. Since $\pi^{\mathrm{R}}(Q)$ is differentiable at $Q = B/w$, we conclude that $\pi^{\mathrm{R}}(Q)$ is unimodal in its entire domain $Q \geq 0$. The value of Q^* then follows from solving the first-order condition $\pi^{\mathrm{R}}_{\mathrm{Q}}(Q) = 0$ and imposing the budget constraint $\alpha\,w\,Q^* \leq B$.

Proof of Proposition 11.3: Let us write the supplier's problem as:

$$\max_{\alpha \in [\tilde{\alpha}(w), 1]} \pi^{\mathrm{S}}(\alpha)$$

For the auxiliary function:

$$\pi^{\mathrm{S}}(\alpha) = \left(1 - \frac{c}{w}\right)\frac{B}{\alpha} - \mathbb{E}\left[\left(D - (1-\alpha)\,\frac{B}{\alpha}\right)^-\right]$$

Let us prove that $\pi^{\mathrm{S}}(\alpha)$ is a unimodal function for $\alpha \in [0, 1]$. Indeed, it is not hard to show that:

$$\pi^{\mathrm{S}}_{\alpha}(\alpha) \triangleq \frac{\mathrm{d}\pi^{\mathrm{S}}(\alpha)}{\mathrm{d}\alpha} = \frac{B}{\alpha^2}\left[F\left(\frac{(1-\alpha)\,B}{\alpha}\right) - \frac{w - c}{w}\right]$$

Therefore, $\pi^{S}(\alpha)$ has a unique extreme value $\hat{\alpha}(w) \in (0, 1)$ (solution of $\pi^{S}_{\alpha}(\alpha) = 0$) given by:

$$\hat{\alpha}(w) = \frac{B}{B + \bar{F}\left(\frac{c}{w}\right)}.$$

In addition:

$$\left.\frac{d^2\pi^{S}(\alpha)}{d\alpha^2}\right|_{\alpha=\hat{\alpha}} = -\frac{B^2}{\alpha^4} f\left(\frac{(1-\alpha)\,B}{\alpha}\right) < 0$$

This prove that $\hat{\alpha}$ is a maximum and so $\pi^{S}(\alpha)$ is unimodal as claimed. As a result the solution to supplier's problem above is given by:

$$\alpha^*(w) = \max\{\hat{\alpha}(w)\ ,\ \tilde{\alpha}(w)\}$$

Finally, using the definition of $\tilde{\alpha}(w)$ and $\tilde{Q}(w)$ is a matter of straightforward manipulations to show that:

$$\tilde{\alpha}(w) = \frac{B}{B + \bar{F}^{-1}\left(\frac{1}{w}\,\bar{F}(\tilde{Q}(w))\right)}.$$

Proof of Proposition 11.4: In the following proof, we use some of the notation introduced in the proof of Proposition 11.3.

By Lemma 11.1, the function $\tilde{Q}(w)$ is decreasing in w. Furthermore, we have $\bar{F}(\tilde{Q}(c)) = c\,\bar{F}(c\,\tilde{Q}(c) - B) \le c$ and $\bar{F}(\tilde{Q}(1)) = 1 \ge c$ (since $\tilde{Q}(1) = 0$ when $B > 0$), and so by continuity we must have that there exists a $\dot{w} \in [c, 1]$ such that $\bar{F}(\tilde{Q}(\dot{w})) = c$ and $\bar{F}(\tilde{Q}(w)) \le c$ for $w \in [c, \dot{w}]$ and $\bar{F}(\tilde{Q}(w)) \ge c$ for $w \in [\dot{w}, 1]$. It follows from Proposition 11.3 and Corollary 11.1 that for all $w \in [c, \dot{w}]$:

$$(\alpha_I(w), Q_I(w)) = \left(\hat{\alpha}(w), \frac{B}{w\,\hat{\alpha}(w)}\right)$$

$$\text{and} \quad \pi^{S}(w) = \left(1 - \frac{c}{w}\right)\frac{B}{\hat{\alpha}(w)} - \mathbb{E}\left[\left(D - (1-\hat{\alpha}(w))\,\frac{B}{\hat{\alpha}(w)}\right)^-\right]$$

Furthermore, from the definition of $\hat{\alpha}(w)$, it follows that $\partial\pi^{S}(w)/\partial\hat{\alpha}(w) = 0$ and so:

$$\frac{d\pi^{S}(w)}{dw} = \frac{\partial\pi^{S}(w)}{\partial\hat{\alpha}(w)}\,\frac{d\hat{\alpha}w}{dw} + \frac{\partial\pi^{S}(w)}{\partial w} = \frac{\partial\pi^{S}(w)}{\partial w} > 0, \quad \text{for all } w \in [c, \dot{w}]$$

We conclude that the optimal w^* belongs to $[\dot{w}, 1]$. In this region, $\alpha_I(w) = \tilde{\alpha}(w)$ and $Q_I(w) = \tilde{Q}(w)$ as a result the supplier's payoff is given by:

$$(w - c)\,\tilde{Q}(w) - \mathbb{E}[(D + B - w\,\tilde{Q}(w))^-], \quad \text{for all } w \in [\dot{w}, 1]$$

To prove the monotonicity of the supplier's payoff on B, let us define the function $w(\tilde{Q}, B)$ as the unique solution of $\bar{F}(\tilde{Q}) = w\,\bar{F}(w\,\tilde{Q} - B)$. In other words, $w(\tilde{Q}, B)$ is the inverse function of $\tilde{Q}(w, B)$, which is guaranteed to exists by Lemma 11.1. Then, for a fixed $\tilde{Q}$, we can show that the supplier's payoff satisfies:

$$\Pi_{\mathrm{I}}^{\mathrm{s}}(\tilde{Q}, B) = \tilde{Q}\,(\bar{F}(\tilde{Q}) - c) + \int_0^{w(\tilde{Q},B)\,\tilde{Q}-B} [D - (w(\tilde{Q}, B)\,\tilde{Q} - B)]\,\mathrm{d}F(D)$$

For notational convenience, in what follows we drop the dependence of $w(\tilde{Q}, B)$ on both argument. Taking partial derivative with respect to B, it follows that:

$$\begin{aligned}
\frac{\partial \Pi_{\mathrm{I}}^{\mathrm{s}}(\tilde{Q}, B)}{\partial B} &= F(w\,\tilde{Q} - B) + \bar{F}(w\,\tilde{Q} - B)\,\tilde{Q}\,\frac{\partial w}{\partial B} \\
&= F(w\,\tilde{Q} - B) - \bar{F}(w\,\tilde{Q} - B)\left[\frac{w\tilde{Q}f(w\tilde{Q} - B)}{\bar{F}(w\,\tilde{Q} - B) - w\,\tilde{Q}f(w\,\tilde{Q} - B)}\right] \\
&= 1 - \frac{\bar{F}(w\,\tilde{Q} - B)}{1 - w\,\tilde{Q}\,h(w\,\tilde{Q} - B)} \\
&\leq 1 - \frac{\bar{F}(w\,\tilde{Q} - B)}{1 - (w\,\tilde{Q} - B)\,h(w\,\tilde{Q} - B)} := 1 - g(w\,\tilde{Q} - B)
\end{aligned}$$

The second equality follows from the definition of $w(\tilde{Q}, B)$. By Assumption 11.1, the function $h(x)$ is increasing and so the function that follows satisfies $g(Q) \geq 1$ for all $Q < \hat{Q}$, where $\hat{Q}$ is the unique root of $Q\,h(Q) = 1$.

$$g(Q) := \frac{\bar{F}(Q)}{1 - Q\,h(Q)}$$

We conclude that:

$$\frac{\partial \Pi_{\mathrm{I}}^{\mathrm{s}}(\tilde{Q}, B)}{\partial B} \leq 0,$$

for all w such that $w\,\tilde{Q} - B < \hat{Q}$. It is not hard to show that $\hat{Q}$ maximizes $Q\,\bar{F}(Q)$. This property together with the fact that $w(\tilde{Q}, B)$ is the unique solution of $\tilde{Q}\,\bar{F}(\tilde{Q}) = w\,\tilde{Q}\,\bar{F}(w\,\tilde{Q} - B)$ imply that for any $\tilde{Q}$, we must have $w\,\tilde{Q} - B < \hat{Q}$. We conclude for all $\tilde{Q}$, which completes the proof, that:

$$\frac{\partial \Pi_{\mathrm{I}}^{\mathrm{s}}(\tilde{Q}, B)}{\partial B} \leq 0$$

Proof of Proposition 11.5: Consider $B \leq \hat{B}$ and recall that $\Pi_{\mathrm{I}}^{\mathrm{s}}(w, B) = (w - c)\,\bar{F}^{-1}(w)$ for $B \geq \bar{w}(B)$. Let us compare the left and right derivatives of $\Pi_{\mathrm{I}}^{\mathrm{s}}(w, B)$ at $w = \bar{w}(B)$. For the left derivative of $\Pi_{\mathrm{I}}^{\mathrm{s}}(w, B)$ we use equation (11.7) at $\bar{w}(B)$ and the fact that $\bar{w}(B)\,\tilde{Q}(\bar{w}(B)) - B = 0$ to get:

$$\left.\frac{\mathrm{d}\Pi_{\mathrm{I}}^{\mathrm{S}}(w, B)}{\mathrm{d}w}\right|_{w\uparrow\bar{w}(B)} = \tilde{Q}(\bar{w}(B)) + (w-c)\left.\frac{\mathrm{d}\tilde{Q}(w)}{\mathrm{d}w}\right|_{w\uparrow\bar{w}(B)}$$

Using the fact that $\tilde{Q}(\bar{w}(B)) = \bar{F}^{-1}(\bar{w}(B))$ and the expression for the derivative of $\tilde{Q}(w)$ with respect to w obtained in the proof of Lemma 11.1, we get that:

$$\begin{aligned}
&\left.\frac{\mathrm{d}\Pi_{\mathrm{I}}^{\mathrm{S}}(w, B)}{\mathrm{d}w}\right|_{w\uparrow\bar{w}(B)} \\
&= \bar{F}^{-1}(\bar{w}(B)) - \frac{\bar{w}(B)-c}{f(\bar{F}^{-1}(\bar{w}(B)))} + \frac{(\bar{w}(B)-c)\,h(0)}{h(\tilde{Q}(\bar{w}(B))}\left[\frac{1-\tilde{Q}(\bar{w}(B))\,h(\tilde{Q}(\bar{w}(B)))}{\bar{w}(B)\,h(0)-h(\tilde{Q}(\bar{w}(B))}\right] \\
&= \left.\frac{\mathrm{d}\Pi_{\mathrm{I}}^{\mathrm{S}}(w, B)}{\mathrm{d}w}\right|_{w\downarrow\bar{w}(B)} + \frac{(\bar{w}(B)-c)\,h(0)}{h(\tilde{Q}(\bar{w}(B))}\left[\frac{1-\tilde{Q}(\bar{w}(B))\,h(\tilde{Q}(\bar{w}(B)))}{\bar{w}(B)\,h(0)-h(\tilde{Q}(\bar{w}(B))}\right] \\
&\leq \left.\frac{\mathrm{d}\Pi_{\mathrm{I}}^{\mathrm{S}}(w, B)}{\mathrm{d}w}\right|_{w\downarrow\bar{w}(B)}
\end{aligned}$$

The (weak) inequality follows from the fact that the term in the square bracket is negative, a fact that can be prove using a similar argument to the one used in the proof of Lemma 11.1.

Now, if $f(0) = 0$ then $h(0) = 0$ and the inequality is an equality. Hence, the supplier's payoff is smooth at $w = \bar{w}(B)$. The result follows from this observation and the following facts: (*i*) $\Pi_{\mathrm{I}}^{\mathrm{S}}(w, B)$ is unimodal in $w \in [\underline{w}(B), \bar{w}(B)$, $(w - c)\,\bar{F}^{-1}(w)$ is unimodal in $w \in [\bar{w}(B), 1]$, and $(w-c)\,\bar{F}^{-1}(w)$ is maximized at $w = w_{\mathrm{E}}$. Therefore, if $B \geq B_{\mathrm{E}}$ then $\bar{w}(B) \leq w_{\mathrm{E}}$ and we must have $w_{\mathrm{I}}(B) = w_{\mathrm{E}}$.

Proof of Lemma 11.2: Let us first suppose (by contradiction) that $\mathbb{E}[\mathcal{L}] < L$ then by Jensen's inequality we get that for all r:

$$\mathbb{E}\left[\min\{\mathcal{L}, L\,(1+r)\}\right] \leq \min\{\mathbb{E}[\mathcal{L}], L\,(1+r)\} < L$$

Therefore, for a loan $(L, \mathcal{L}, r)$ to be feasible in a competitive financial market we must have $\mathbb{E}[\mathcal{L}] \geq L$.

To prove the sufficient condition, let us define $G(r) = \mathbb{E}\left[\min\{\mathcal{L}, L\,(1+r)\}\right] - L$, which is a continuous function of r. Next note that if $\mathbb{E}[\mathcal{L}] \geq L$ then $G(0) \leq 0$ and $G(\infty) \geq 0$. Therefore, by continuity there exists and $r \geq 0$ such that $G(r) = 0$.

REFERENCES

Babich V., Aydm G., Brunet Y., Keppo J., Saigal R. 2008. Risk, financing and the optimal number of suppliers. Technical report, Ann Arbor, Michigan: Industrial and Operations Engineering, University of Michigan.

Babich V., Sobel M. 2004. Pre-IPO operational and financial decisions. *Management Science*, **50**:935–948.

Boyabatli O., Toktay L. 2004. Operational hedging: A review with discussion. Working paper. Fontainebleau, France: INSEAD.

Boyabatli O., Toktay L. 2010. Stochastic capacity investment and flexible versus dedicated technology choice in imperfect capital markets. Working paper, Singapore: Singapore Management University.

Buzacott J., Zhang R. Q. 2004. Inventory management with asset-based financing. *Management Science*, **50**(9):1274–1292.

Cachon G. 2003. Chapter 6. Supply chain coordination with contracts. In: de Kok, A. and Graves S., editors. *Supply Chain Management: Design, Coordination and Operation,* vol. 11 of *Handbooks in Operations Research and Management Science.* Amsterdam, The Netherlands: Elsevier.

Cachon G., Zipkin P. 1999. Competitive and cooperative inventory policies in a two-stage supply chain. *Management Science*, **45**:936–953.

Caldentey R. and Haugh M. 2006. Optimal control and hedging of operations in the presence of financial markets. *Mathematics of Operations Research* **31**(2): 285–304.

Caldentey R., Haugh M. 2009. Supply contracts with financial hedging. *Operartions Research*, **57**(1):47–65.

Chod J., Rudi N., Mieghem J. V. 2009. Operational flexibility and financial hedging: Complements or substitutes? Working paper, Boston: Boston College, available at SSRN: http://ssrn.com/abstract=1385438.

Dada M., Hu Q. 2008. Financing the newsvendor inventory. *Operations Research Lettersss* **36**(5).

de Kok A., Graves S., editors. 2003. *Supply Chain Management: Design, Coordination and Operation*, vol. 11 of *Handbooks in Operations Research and Management Science*. Amsterdam, The Netherlands: Elsevier.

Ding Q., Dong L., Kouvelis P. 2007. On the integration of production and financial hedging decisions in global markets. *Operations Research*, **55**(3):470–489.

Dong L., Kouvelis P., Su P. 2006. Operational hedging strategies and competitive exposure to exchange rates. Working paper, St. Louis. Missouri: Olin Business School, Washington University.

Gaur V., Seshadri S. 2005. Hedging inventory risk through market instruments. *Manufacturing and Service Operations Management*, **7**(2):103–120.

Hadley G., Whitin T. 1963. *Analysis of Inventory Systems*. Toronto, Canada: Prentice Hall.

Hu Q., Sobel M. 2005. Capital structure and inventory management. Working paper. Cleveland, Ohio: Case Western Reserve University.

Kouvelis P., Zhao W. 2008. Financing the newsvendor: Supplier vs. bank, optimal rates, and alternative schemes. Working paper. St. Louis, Missouri: Olin Business School, Washington University.

Lariviere M. 2006. A note on probability distributions with increasing generalized failure rates. *Operations Research*, **54**(3):602–604.

Lariviere M., Porteus E. 2001. Selling to the newsvendor: An analysis of price-only contracts. *Manufacturing Service Operations Management*, **3**(4):293–305.

Milne R. 2009. Tactics that deliver in a 'war for cash'. *Financial Times*, January.

Simchi-Levi D., Wu S. D., Shen Z., editors. 2004. *Handbook of Quantitative Supply Chain Analysis: Modeling in the eBusiness Era*. New York: Springer Science.

Spengler J. 1950. Vertical integration and antitrust policy. *Journal of Political Economy* **58**:347–352.

Stiglitz J. and Weiss J. 1981. Credit rationing in markets with imperfect information. *American Economic Review* **71**(3):393–410.

Xu X., Birge J. 2004. Joint production and financing decisions: modeling and analysis. Working paper. Chicago, Illinois: Booth School of Business, University of Chicago. Available at SSRN: http://ssrn.com/abstract=652562.

CHAPTER TWELVE

Production/Inventory Management and Capital Structure

QIAOHAI (JOICE) HU, LODE LI, AND MATTHEW J. SOBEL

12.1 Operations and Finance

Businesses manage flows of material and cash, and large firms decentralize these tasks into separate functional responsibilities in spite of the interactions; material needs capital and sales of goods contributes to cash reserves. Management research and education mimics this dichotomy into operations and finance. The research and pedagogical literatures on production/inventory management focus on operations analyses without being concerned directly with finance, and the analogous finance literatures focus on financial decisions while suppressing operational details. We show that professional practice that is consistent with these dichotomies diminishes the value of firms.

Acknowledgments: The author is grateful for the thoughtful and constructive suggestions by the three anonymous reviewers.

Handbook of Integrated Risk Management in Global Supply Chains, First Edition. Edited by Panos Kouvelis, Lingxiu Dong, Onur Boyabatli, and Rong Li.

It is tempting to separate operations and finance. It is easier to analyze these functions separately than when they are combined, and there are practical benefits due to specialization and a restricted focus. When the coordination costs were high, for example prior to the widespread implementation of enterprise-wide information systems, there was much to gain from such decentralization or specialization. However, information technology has advanced so fast that in many firms it is technically as well as economically feasible to coordinate financial and operational decisions. Thus we confront a practical question: When is the separation of operations and finance a good enough approximation of reality, and when should the functions be coordinated because the interactions are too important to ignore?

The Modigliani-Miller Theorem (Modigliani and Miller 1958) forms the basis for modern thinking about capital structure and provides a partial answer to the above question. The basic theorem states that, in the absence of taxes, bankruptcy costs, and asymmetric information, and in an efficient market, the value of a firm is unaffected by how that firm is financed. It does not matter if the firm's capital is raised by issuing stock or incurring debt. The dividend policy does not affect the firm's value. This capital structure irrelevance principle justifies the standard inventory-management approach of ignoring financial constraints and separating the operational and financial decisions. In reality, where market frictions such as taxes and bankruptcy costs conspicuously violate the perfect market assumptions underlying the Modigliani-Miller Theorem, the separation of the operations and finance functions may in fact hamper a firm's performance. See Section 12.4.3 for numerical examples of performance deterioration, and see Section 12.8 for further discussion of the relevant literature. It is important to study the coordination of the two types of decisions to address questions such as: What is the opportunity cost of separating operational and financial decisions? What does coordination entail and what is its value?

This chapter summarizes the findings in Li, Shubik and Sobel (1997, revised 2008) and Hu and Sobel (2005), which focus on the following issues regarding the coordination of operational decisions and financial decisions and on how such efforts affect the firm's capital structure:

- What is the opportunity cost of separating operational and financial decisions? In other words, what is the value of coordinating operations and finance?
- Under what circumstances is this value of functional coordination high? Under what circumstances is the cost of functional specialization negligible?
- When the value of coordination is large, what does coordination entail? How do coordinated operational decisions differ from their counterparts without financial considerations? Similarly, how do coordinated financial decisions differ?
- What are the relationships between a firm's long-term capital structure decisions and its short-term operational and financial decisions? For example,

how does the level of long-term debt affect the firm's operating decisions regarding inventory, working capital and short-term loans?

- How does a firm determine an optimal capital structure in coordination with its short-term operational and financial decisions?

This chapter answer these questions when the consequences of default are consistent with Chapter 11 of the U.S. Bankruptcy Code (i.e., bankruptcy precipitates a costly reorganization of the firm). This is the most common form of business bankruptcy in the U.S. A less common but perhaps more sensational form of bankruptcy is via Chapter 7 in which the owners lose any claim to future earnings of the firm. Some of the conclusions are invariant with respect to the form of bankruptcy; see Section 12.6.1 and Hu et al. (2010).

The rest of the chapter is structured as follows. Section 12.2 specifies a dynamic stochastic model in which the firm's inventory and financial decisions interact and are coordinated in the presence of demand uncertainty, a financial constraint, and a default risk. Section 12.3 analyzes the firm's coordinated production, dividend and short-term borrowing decisions, and establishes key properties of a dynamic programming representation of the optimization problem. In particular, we reduce the dimensionality of the dynamic program and show that the reduced dynamic program has a concave structure and a myopic optimal policy which is consistent with the "pecking-order" principle of corporate finance. Section 12.4 considers an important case in which the default penalty and the gross profit from sales (net of inventory costs) are piecewise linear functions. The optimal inventory, working capital, and dividend policies are explicitly characterized in this case. We then compare the operational decisions stemming from financial considerations with their corresponding values, which optimize the expected present value (EPV) of the profits (*not* dividends), without cash constraints. We also present numerical examples to illustrate (a) how the coordinated decisions differ from their decentralized counterparts, and (b) the value of such coordination. Section 12.5 embeds the firm's long-term capital structure decisions, equity, and long-term debt, in the basic model and investigates the interaction between capital structure and production/inventory decisions. Section 12.5.1 investigates the impact of long-term debt on the coordinated short-term operational and financial decisions. Section 12.5.2 contrasts the results in Section 12.5.1 for coordinated and decentralized operations decisions. Section 12.5.3 examines the firm's optimal capital structure decision. Section 12.6 discusses the extensions and variations of the basic model, including "wipeout" bankruptcy and the effects of precluding capital subscriptions. We show that the key results from the basic model remain valid (qualitatively) except the optimal policy with non-negative dividends is no longer myopic. Concluding remarks and future research directions are in Section 12.7. Most of the research on the fusion of operations and finance is recent in origin, and in Section 12.8 we comment on the work that is most relevant to this chapter.

12.2 The Model

We consider a discrete-time model of a firm living in perpetuity that decides each period, prior to the occurrence of the period's demand, the number of units to produce, the amount of money to borrow, and the amount of a dividend to declare. Although actual firms may have recourse to several means of financing, until Section 12.5 we assume that the firm can obtain only short-term loans. The specifics of the model are as follows.

Let x_n and w_n, respectively, denote the amounts of inventory (net of cumulative unsatisfied demand, if any) and of retained earnings at the beginning of period n $(n = 1, 2, \cdots)$. In each period n, the firm makes three decisions: the amount of short-term loan borrowed at the start of the period and repaid at the end of the period, b_n; the quantity of goods procured/produced, z_n; and the amount of dividend issued net of capital subscription, v_n (i.e., v_n is the dividend paid if $v_n > 0$ and its absolute value is the amount of a capital subscription if $v_n < 0$). The interest rate on short-term loans is ρ and ρb_n is interest paid at the beginning of period n. The results would not significantly change if we assumed that interest were paid at the end of the period. The only exogenous risk considered in our model is demand uncertainty, and that induces a default risk. Let D_n denote the new demand in period n. Assume that $D_1, D_2, \ldots$ are independent, identically distributed, and non-negative, and let F and f denote the respective cumulative and density distribution functions of D, which has the same distribution as D_n. We assume that demand is backordered if it exceeds the amount of goods on hand.

Let c be the unit procurement cost, r be the unit selling price, and γ be the unit backorder cost. So r is the revenue received in the period when a unit of new demand is met and $r - \gamma$ is the revenue received when a backorder is met. We assume $r > c$.

Under the assumption that the ordered quantity z_n is received sufficiently quickly to satisfy demand in period n, the total amount of product available to satisfy the current demand is $y_n = x_n + z_n$. The sales revenue from new demand net of inventory cost in period n, denoted $g(y_n, D_n)$, is a function of total supply and demand. An example of such function with a linear holding cost is $g(y, d) = r\min\{y, d\} - h(y - d)^+ = ry - (r + h)(y - d)^+$ where d and h, respectively, denote realization of demand and unit inventory cost. Because of the demand uncertainty, the retained earnings for the next period may become negative (when demand is low and inventory is high). If $w_n < 0$, a default penalty, $p(w_n)$, will be imposed but it is convenient to define $p(\cdot)$ as a function on $\Re$.

Although many of our results depend on $p(\cdot)$ being a convex decreasing function, the most important results do not depend on this assumption and we make no assumptions now regarding $p(\cdot)$. Although we have called $p(\cdot)$ a "default" penalty, it encompasses three related phenomena. First, if the retained earnings go negative and the firm cannot refinance, this could be the terminal bankruptcy costs. Second, if the firm can refinance, it can do so by borrowing externally and paying legal fees and other financing costs. Third, it can raise money by subscriptions from current stockholders. Our current model illustrates the latter two choices (the terminal bankruptcy case will be dealt with in Section 12.6). Thus, the default penalty can be thought of as the total amount of corporate losses, including the utility losses of stockholders, due to an insolvency.

The sequence of decisions and events n is the same in each period n and is as follows. The first event is the computation of the amount of retained earnings (w_n) and the current inventory level (x_n). The default penalty ($p(w_n)$) is paid if the amount of retained earnings is negative. Then there is a selection of the amounts to borrow (b_n) and produce (z_n), and the amount of a dividend (v_n) to declare. Having made those choices, the firm pays the dividend and loan interest (v_n and ρb_n), it implements the production decision at a cost of cz_n, and it satisfies the backordered demand with a revenue of $(r - \gamma)x_n^-$ where the notation x_n^- denotes $(-x_n)^+ = \max\{-x_n, 0\}$. Then the new demand ($D_n$) is realized and the firm receives the corresponding sales revenue net of inventory costs ($g(y_n, D_n)$). Finally, the firm repays the principal of the loan (b_n).

We assume that the revenue net of inventory-related costs in period n is received at the end of the period and is not available to pay the dividend at the beginning of the period. Similarly, we assume that inventory-related costs depend on the residual inventory after demand occurs. So the end-of-period inventory costs in period n do not affect the availability of funds to pay the dividend at the beginning of period n.

We do not micro-model the consequences if the firm is unable to repay b_n at the end of period n, but our model encompasses a variety of possibilities. Here is one example. Suppose that each period, if the firm has not completely repaid prior periods' loans including interest on unpaid portions, it is charged a (presumably high) interest rate $\mathcal{I}$. Then the bankruptcy penalty $p(w_n)$ would include $\mathcal{I} \cdot (w_n)^-$ because $(w_n)^-$ is precisely the cumulative amount of prior periods' unpaid short-term loans plus interest on unpaid loans.

It simplifies the analysis to define new decision variables:

$$y_n \equiv x_n + z_n \tag{12.1}$$

$$s_n \equiv w_n + (r - \gamma)x_n^- - p(w_n) - v_n - cz_n - \rho b_n \tag{12.2}$$

Under the assumption that ordered goods are delivered immediately, y_n is the amount of goods that is available to satisfy demand in period n. We assume that $y_n \geq 0$, so $z_n \geq -x_n$. If demand exceeds the supply of goods, we assume that the excess demand is backordered and that the backordered demand will be satisfied before meeting new demand in the next period. That means, if $x_n < 0$, that the production quantity, z_n, is at least as great as the backordered demand, $-x_n$. This assumption is reasonable because the margin on the backordered demand is positive and risk-free.

The amount of internally generated working capital available to finance short-term decisions is s_n. So s_n is the amount of working capital after paying the dividend, loan interest, and production cost, and receiving the revenue on backordered demand, but before making the loan, receiving the sales revenue, or paying the inventory costs. Therefore, b_n+s_n is period n's total available working capital, after all decisions are made and implemented but before demand and sale revenues are realized. Revenues and inventory costs are realized by the end of the period, and the loan principle is straightforwardly repaid if $b_n + s_n + g(y, D_n) \geq b_n$. Otherwise, there is a delay in repayment until there is enough retained earnings to repay it, and the default penalty $p(w_{n+1})$ is levied. Note that we do not include either interest or discounting of the loan principle when repayment is delayed. The reason is that both can be seamlessly incorporated in the default penalty function.

The equations that balance the material flow and cash flow are $x_{n+1} = x_n + z_n - D_n$ and:

$$w_{n+1} = w_n + (r - \gamma)x_n^- - p(w_n) - v_n - cz_n + g(y_n, D_n) - \rho b_n. \quad (12.3)$$

With the convenient notations (12.1) and (12.2), the two equations become:

$$x_{n+1} = y_n - D_n \quad (12.4)$$

$$w_{n+1} = s_n + g(y_n, D_n) \quad (12.5)$$

If excess demand were lost instead of backordered, (12.4) would be replaced by $x_{n+1} = (y_n - D_n)^+$ and all of the paper's qualitative results would be preserved (see Hu and Sobel 2005).

Under the assumption that the loan and production quantities are non-negative, we constrain:

$$b_n \geq 0 \quad \text{and} \quad z_n \geq 0 \quad (12.6)$$

The inequality $w_n + (r - \gamma)x_n^- + (1 - \rho)b_n \geq p(w_n) + v_n + cz_n$ is the liquidity constraint, and it is equivalent to:

$$b_n + s_n \geq 0 \quad (12.7)$$

This inequality ensures that the sum of retained earnings plus the loan proceeds is at least as great as the expenditures in period n. However, (12.7) forces v_n to be negative if $w_n + (r - \gamma)x_n^- + (1 - \rho)b_n < p(w_n) + cz_n$. In Section 12.6.2 we constrain dividends to be non-negative (i.e., $v_n \geq 0$, and analyze the consequences).

Given x_n and w_n, it follows from (12.1) and (12.2) that the decision variables in period n, namely z_n, v_n, and b_n, are equivalent to y_n, s_n, and b_n. It is convenient to use the latter specification as the decision variables. Let β denote the single period discount factor ($0 < \beta < 1$), for $n = 1, 2, \ldots$. Borrowing would not occur unless the interest rate is less than the opportunity cost of capital, so let $\rho < 1/\beta - 1$. Let H_n denote the elapsed history $H_n \equiv (x_1, w_1, b_1, s_1, y_1, D_1, \ldots, x_{n-1}, w_{n-1}, b_{n-1}, s_{n-1}, y_{n-1}, D_{n-1}, x_n, w_n)$, and let B denote the present value of the dividends net of capital subscriptions. So:

$$B = \sum_{n=1}^{\infty} \beta^{n-1} v_n \quad (12.8)$$

Although the stockholders may be diverse, we treat their preferences as representing the "firm's" objective. A *policy* is a rule that, for each n, chooses y_n, s_n, and b_n as a function of H_n. That is, a policy is a nonanticipative rule for choosing $y_1, s_1, b_1, y_2, s_2, b_2, \ldots$. A policy is *optimal* if it maximizes $E[B|x_1 = x, w_1 = w]$ for each $(x, w) \in \Re^2$. Our goal is to characterize an optimal policy.

All models, including ours, suppress many details to facilitate analyses and secure insights. The inclusion of some of those details in our model, such as shareholder income taxes, would complicate the analysis and affect the results. However, several additional details would *not* significantly affect the results but would obscure the exposition. First, the dividend decision could be made less frequently than the borrowing decision, and the borrowing decision could be made less often than the production quantity decision. Second, the right side of (12.3) could include a credit for interest earned on retained earnings.

Third, instead of the borrowing interest rate, ρ, being constant, the rate in period n could be a random variable whose distribution depends on state and decision variables. Similarly, the discount factor in period n could be the n-fold product of random variables whose distributions depend on state and decision variables. See Babich and Sobel (2004) for an analysis of a model with such dependencies. Similarly, loan interest could be paid at the end of the period instead of prepaid at the beginning of the period. Our results would remain valid except in Section 12.4.2 where straightforward changes would be necessary. Fourth, the representative shareholder's attitude towards risk could be included in the model with an exponential utility function. This would preserve the paper's qualitative results but slightly complicate the formulas in Section 12.4.

12.3 Structural Properties of an Optimal Policy

Li et al. (1997) show that the problem has a more compact structure and a simpler optimal policy than might first appear. Although there seem to be two state variables (x_n and w_n) and three decision variables (b_n, s_n, y_n), the problem can be reduced to one with only one state variable (x_n) and two decision variables (s_n and y_n) and its solution can be specified almost explicitly.

Using the flow balance equations (12.4) and (12.5) we can substitute the dividend and production variables v_n and z_n with the new decision variables y_n and s_n defined in (12.1) and (12.2) and rewrite the objective function (12.8) as:

$$
\begin{aligned}
B = {} & (r-\gamma)x_1^- + cx_1 + w_1 - p(w_1) + (r-\gamma-c)\sum_{n=1}^{\infty}\beta^n D_n \\
& + \sum_{n=1}^{\infty}\beta^{n-1}[-(1-\beta)(s_n + cy_n) - \beta(r-\gamma)y_n \\
& + \beta(r-\gamma)(y_n - D_n)^+ + \beta g(y_n, D_n) - \beta p(s_n + g(y_n, D_n)) - \rho b_n]
\end{aligned}
\tag{12.9}
$$

It is convenient to aggregate the terms in the second and third lines of (12.9) as follows. For $(b, y, s) \in \Re^3$, let:

$$K(b, s, y) = -(1-\beta)s - [\beta(r-\gamma) + (1-\beta)c]y + \beta E[(r-\gamma)(y-D)^+ + g(y, D) - p(s + g(y, D))] - \rho b \tag{12.10}$$

12.3.1 REDUCTION OF DIMENSIONALITY

An inventory and financial policy maximizes $E[B|H_1]$ if and only if it maximizes $E[B|H_1] - (r-\gamma)x_1^- + cx_1 + w_1 - p(w_1) + (\beta(r-\gamma-c)/(1-\beta))E[D]$. The latter problem, by using (12.9) and (12.10), is equivalent to optimizing the following criterion on which we focus henceforth:

$$E\left[\sum_{n=1}^{\infty} \beta^{n-1} K(b_n, s_n, y_n)\right] \tag{12.11}$$

The constraints on the decision variables are:

$$y_n \geq 0, \quad y_n \geq x_n, \quad b_n + s_n \geq 0, \quad \text{and} \quad b_n \geq 0 \tag{12.12}$$

The maximization of (12.11) subject to (12.12) is equivalent to the optimization of the expected value of (12.8) subject to (12.6) and (12.7). Although the state in a dynamic program for the latter problem has a pair of scalars, x_n and w_n, we observe in (12.12) that w_n is neither an argument of K nor does it constrain the decision variables y_n, b_n, and s_n (due to the implicit assumption of unlimited credit). Therefore, w_n is a redundant state variable and the optimization of the latter problem is equivalent to the optimization of the former problem *which has only a single scalar state variable*. This reduces the computational work to get numerical solutions, facilitates the characterization of an optimal policy, and is the basis of the following statement.

Proposition 12.1 *The optimization of the expected value of (12.8) subject to (12.6) and (12.7) corresponds to the following dynamic program:*

$$\psi(x) = \sup_{b,s,y} \{J(b, s, y) : y \geq 0, y \geq x, b \geq 0, b + s \geq 0\} \tag{12.13}$$

$$J(b, s, y) = K(b, s, y) + \beta E[\psi(y - D)] \tag{12.14}$$

The infinite-horizon dynamic program (12.13) and (12.14) corresponds to the finite-horizon recursion $\psi_{N+1}(\cdot) \equiv 0$ and:

$$\psi_n(x) = \max_{b,s,y} \{J_n(b, s, y) : y \geq 0, y \geq x, b \geq 0, b + s \geq 0\} \tag{12.15}$$

$$J_n(b, s, y) = K(b, s, y) + \beta E[\psi_{n+1}(y - D)] \tag{12.16}$$

for each $n = 1, 2, \ldots, N$ and $x \in \Re$. In (12.15), let $b_n(x)$, $s_n(x)$, and $y_n(x)$ be optimal values of b, s, and y, respectively. We use the finite-horizon approximation ψ_n to deduce properties of (12.13) and (12.14).

12.3.2 CONCAVITY AND MONOTONICITY

The following result gives conditions which imply that the marginal value of inventory increases as the planning horizon lengthens. This is an intuitive property of many dynamic concave resource allocation models (Mendelssohn and Sobel 1980). If the horizon is longer, then there is greater opportunity to make productive use of an additional unit of resource. Here and throughout the chapter, "decrease" and "increase" are used in the weak sense.

Proposition 12.2 *Suppose that* $p(\cdot)$ *is a decreasing convex function on* $\Re$ *and* $g(\cdot, d)$ *is a concave function on* $\Re$ *for each* $d \in \Re_+$.

1. *The value function in (12.15),* $\psi_n(\cdot)$, *is a concave function on* $\Re$ *and* $J_n(\cdot, \cdot, \cdot)$ *is a concave function on* $\Re^3$ *for each n.*
2. *Let* $\psi'_n(x)$ *be the right-hand derivative of* $\psi_n(x)$*; then* $\psi'_n(x) \leq \psi'_{n+1}(x)$ *for each n and x.*
3. *For each n,* $y_n(x)$ *is increasing and* $z_n(x)$ $(y_n(x) - x)$ *is decreasing with respect to* $x \in \Re$. *If* $p(\cdot)$ *is decreasing on* $\Re$ *then* $v_n(x, \cdot)$ *is increasing on* $\Re$ *for each* $x \in \Re$.

The second part of Proposition 12.2 asserts that, if the horizon is longer, then there is greater opportunity to make productive use of an additional unit of resource. The third part shows that the monotonicity of the target inventory level and the production/purchase quantity still hold for our model, which is shared by optimal policies in many production/inventory models. It is consistent with current inventory and additional goods that are procured or produced being substitutes. The dividend is monotone due to the fact that an increase in "cash on hand," w, loosens the liquidity constraint. An increment in w permits at least a small increase in the dividend while allowing the selection of the order-up-to inventory level (y) and residual retained earnings (s) that were optimal at the original level of w.

We define:

$$G(y) = \sup_{b,s}\{K(b, s, y) : b \geq 0,\, b + s \geq 0\} \tag{12.17}$$

We also rewrite (12.13) and (12.14) as:

$$\psi(x) = \sup_{y}\{G(y) + \beta E[\psi(y - D)] : y \geq 0, y \geq x, \} \tag{12.18}$$

This is the dynamic program for the dynamic newsvendor model and, as a result, in the current model there is an optimal base-stock level policy for inventory replenishment. Indeed, a myopic policy is optimal.

We do *not* need the convexity and concavity assumptions in Proposition 12.2 for any of the remaining results in this section.

12.3.3 A MYOPIC OPTIMUM

Let $(b, s, y) = (b^*, s^*, y^*)$ maximize $K(b, s, y)$ subject to $b \geq 0$, $b + s \geq 0$, and $y \geq 0$:

$$K(b^*, s^*, y^*) = \sup_{b,s,y}\{K(b, s, y) : b \geq 0, b + s \geq 0, y \geq 0\} \qquad (12.19)$$

Under reasonable assumptions (see Proposition 12.2), this numerically easy nonlinear programming problem is a concave maximization problem with three variables and three polyhedral constraints. Denote the maximal EPV of (12.11) with L. Then:

$$L \leq \frac{K(b^*, s^*, y^*)}{1 - \beta}$$

The upper bound is achieved if $(b_n, s_n, y_n) = (b^*, s^*, y^*)$ is feasible in (12.11) for all n. A standard initial condition yields feasibility. If $x_n \leq y_n^*$ for some n, then $(b_n, s_n, y_n) = (b^*, s^*, y^*)$ is feasible because $b^* \geq 0$, $b^* + s^* \geq 0$, and $y^* \geq 0$ from (12.19). If $y_n = y^*$, then $x_{n+1} = y_n - D_n = y^* - D_n \leq y^*$; so $y_{n+1} = y^*$ is feasible. Therefore, $x_1 \leq y^*$ permits $(b_n, s_n, y_n) = (b^*, s^*, y^*)$ for all n. This series of steps is the basis of the next proposition, which is most useful if $x_1 \leq y^*$.

Proposition 12.3 *If $x_k \leq y^*$ for some k then $(b_n, s_n, y_n) = (b^*, s^*, y^*)$ for all $n \geq k$ is optimal.*

If $x_1 \leq y^*$ then it is feasible, hence optimal, to choose $b_n = b^*$, $s_n = s^*$, and $y_n = y^*$ for all n. There are three consequences. First, if the initial inventory of goods is not too high, then three scalars determine an optimal decision rule. Second, the simplicity of time-invariant scalars instead of the complexity of time-varying functions clarifies the connection (in later sections of this chapter) between dividends and costs. Third, let $b_n(x)$, $s_n(x)$, and $y_n(x)$ denote optimal selections of the decision variables b, s, and y in the finite-horizon dynamic program (12.15) and (12.16). If $x_1 \leq y^*$, then $b_n(x)$, $s_n(x)$, and $y_n(x)$ are invariant with respect to n. We denote their common values $b(x)$, $s(x)$, and $y(x)$ and observe that their time invariance implies that they are optimal in the infinite-horizon problem too.

Another result is that the sequence of production quantities and dividends is a sequence of independent and identically distributed random vectors.

Corollary 12.1 *For all n if $b_n = b^*$, $s_n = s^*$, and $y_n = y^*$, then $(v_2, z_2), (v_3, z_3), \ldots$ are independent and identically distributed random vectors with the same joint distribution as the random vector*

$$((r - \gamma)(y^* - D)^- + g(y^*, D) - cD - p(s^* + g(y^*, D)) - \rho b^*, D)$$

The myopic base-stock policy directs the firm to raise the amount of product stock to a target level y^* and issue dividends and borrow short-term to move the cash stock to a target level s^*. Therefore, production exactly offsets the consumption of goods, and the dividend equals the revenue net of production/inventory costs, default penalty and loan interest payment. Notice that the dividends fluctuate with the demand, because the latter affects the pre-dividend net profit. Thus, dividends vary in order for the firm to maintain a constant level of cash stock.

In a dynamic newsvendor model without a capacity constraint, it is well known that current production equals the most recent demand. However, the joint distribution of production and dividend, and the probability distribution of the dividend are new results. It follows from Corollary 12.1 that each period's dividend should be a reflection of the previous period's actual financial performance (*not* its expected value), and that performance depends on operational decisions (y^*), operations structure ($g(\cdot, \cdot)$), market characteristics (r), market events (D), and financial characteristics ($p(\cdot)$ and ρ).

12.3.4 PECKING ORDER OPTIMALITY

It is optimal to borrow as little as possible while satisfying the liquidity constraint. That is, if $s(x) \geq 0$, then $b(x) = 0$; if $s(x) < 0$, then borrow just enough to raise the level to 0, namely, $b(x) = -s(x)$. This is consistent with the well-known "pecking order" in finance, which advises a firm to borrow externally only after it has resorted to internal equity. Here is the formal statement.

Proposition 12.4 *In (12.13), $b(x) = s(x)^-$ is optimal for all $x \in \Re$.*

12.4 Characterization of the Optimal Policy

Here we detail the myopic optimal policy and specify it explicitly when the default penalty and the gross profit from sales (net of inventory costs) are piecewise linear functions as follows:

$$p(x) = (-\theta x)^+ \tag{12.20}$$

$$g(y, D) = r\min\{y, D\} - h(y - D)^+$$

We exploit the fact that the second equation can be written as:

$$g(y, D) = ry - (r + h)(y - D)^+ \tag{12.21}$$

The parameter $\theta > 0$ is the unit cost of a default. In this section we exploit the facts that $p(\cdot)$ is decreasing and convex, and $g(\cdot, d)$ is concave. First, we use the piecewise linearity to specialize transformed single-period payoff $K(\cdot, \cdot, \cdot)$ defined in (12.10), and characterize the optimal myopic policies. Second, we compare the optimal operating decisions that are determined with financial considerations to those that are determined without them. Third, we give a numerical example.

The transformed single-period payoff (12.10) becomes:

$$\begin{aligned} K(b,s,y) &= -(1-\beta)s + [\beta\gamma - (1-\beta)c]y - \rho b \\ &\quad - \beta E[(h+\gamma)(y-D)^+ + \theta\left(s+ry-(r+h)(y-D)^+\right)^-] \quad \text{and} \\ K(s^-,s,y) &= -(1-\beta)s + [\beta\gamma - (1-\beta)c]y - \rho s^- \\ &\quad - \beta E[(h+\gamma)(y-D)^+ + \theta\left(s+ry-(r+h)(y-D)^+\right)^-] \end{aligned} \tag{12.22}$$

which is concave and continuous by Proposition 12.4. Then, we can expand the expected default penalty in (12.22) to be

$$E[p_1\left(s+ry-(r+h)(y-D)^+\right)^-] = \begin{cases} \theta\int_0^{(hy-s)/(r+h)}(hy-s-(r+h)x)f(x)dx & \text{if } s+ry>0 \\ \theta(r+h)\int_0^y (y-x)f(x)dx & \text{if } s+ry=0 \\ \theta(-s-ry+(r+h)\int_0^y(y-x)f(x)dx) & \text{if } s+ry<0 \end{cases} \tag{12.23}$$

If $s_n = s$ and $y_n = y$, let $\Gamma(s,y)$ be the set of values of demand D_n which do *not* precipitate default in period $n+1$ and $q(s,y)$ be the probability of no default in period $n+1$. That is, $\Gamma(s,y) = \{d : s + g(y,d) > 0\} = \{d : s + ry - (r+h)(y-D)^+ > 0\}$ and $q(s,y) = P\{D \in \Gamma(s,y)\}$. If demand were to match supply exactly, the resulting level of retained earnings would be $s + ry$. That is the maximum amount of earnings the firm can achieve, and each unit of inventory would reduce it by $r+h$. So if $s + ry < 0$, then the firm will default with certainty regardless of the realization of demand; that is, $\Gamma(s,y) = \emptyset$ and $q(s,y) = 0$. If $s + ry = 0$, then any positive inventory will precipitate default; so default would not occur if and only if supply were to match demand perfectly. In this case, $\Gamma(s,y) = \{y\}$. Finally, if $s + ry > 0$, then the firm will remain solvent only if demand is not too low; that is, if the demand is greater than $(hy-s)/(r+h)$. In this case, $\Gamma(s,y) = \{D : D > (hy-s)/(r+h)\}$ and $q(s,y) = 1 - F[(hy-s)/(r+h)]$. If demand is lower than the lower limit in $\Gamma(s,y)$, then insufficient revenue and high holding costs induce default.

In the next subsection, we consider the whole problem in which the firm may choose to borrow or not, namely, $b \geq 0$. See Section 12.4.1 in Li et al. (2008) for additional results when the firm elects not to borrow (i.e., $b_n = 0$ for all n).

12.4.1 BASE-STOCK POLICIES FOR CASH AND GOODS

In this subsection, we solve:

$$K(b^*,s^*,y^*) = \sup\{K(b,s,y) : b \geq 0,\ b+s \geq 0,\ y \geq 0\}. \tag{12.24}$$

In order to specify the impacts on dividends of adjusting the cash stock level or the product stock level, we differentiate K with respect to s and y. It follows from

(12.22) and (12.23) that:

$$\frac{\partial K}{\partial s} = -(1-\beta) + \begin{cases} \beta\theta F[(hy-s)/(r+h)] & \text{if } s+ry > 0 \\ \beta\theta & \text{if } s+ry < 0 \end{cases} \tag{12.25}$$

and

$$\frac{\partial K}{\partial y} = \beta\gamma - (1-\beta)c - \beta(\gamma+h)F(y) + \begin{cases} -\beta\theta h F[(hy-s)/(r+h)] & \text{if } s+ry > 0 \\ \beta\theta(r-(r+h)F(y)) & \text{if } s+ry < 0 \end{cases} \tag{12.26}$$

Recall that s is the level of cash stock after the firm implements its decisions but before sales revenue and inventory costs are realized. The partial derivatives are intuitive effects on dividends of adjusting the levels of the cash stock, s or the product stock, y. The key trade-off is between the current dividend and the future retained earnings level which will then affect the probability of default. If $s + ry > 0$, then the firm will not default if demand is higher than the default limit, namely $D > (hy - s)/(r + h)$. On the other hand, $s + ry < 0$ represents a situation when the firm will default with certainty. The effects of changing the base-stock level y on the objective are different in the two cases. If we increase the cash stock level, s, by one dollar, the current dividend will decrease by one dollar while next period's retained earnings will increase by one dollar and, hence, the cost of default will go down by θ dollars in the event of a default (with probability $F[(hy - s)/(r + h)]$ if $s + ry \geq 0$ and 1 if $s + ry < 0$). Therefore, equation (12.25) follows.

The effects of increasing y on dividends and retained earnings are indirect via revenues and costs. If y rises by one unit, the production cost increases by c dollars and, hence, the current dividend goes down by c dollars. In contrast, if the unit would be backordered (i.e., demand would exceed supply, with probability $1 - F(y)$), then next period's retained earnings would increase by γ dollars. It will decrease by h dollars if the extra unit produced would not be sold and, therefore, held in inventory for one period (because demand is less than supply, with probability $F(y)$). Furthermore, in the event of default, the extra holding cost will increase the default cost by θh if the firm defaults due to low demand (with probability $F[(hy - s)/(r + h)]$). However, an increase in y has an additional effect on the default cost when $s + ry < 0$. Since the firm will default regardless of demand's realization, one more unit of supply will generate r more dollars of sales and, hence, reduce the default penalty by θr dollars as long as the unit can be sold (with a probability of $1 - F(y)$). On the other hand, one more unit of supply will increase the default cost by θh dollars when the unit cannot be sold (with probability of $F(y)$). Equation (12.26) summarizes the above trade-offs with respect to a marginal change of y.

Define:

$$y_0^* = F^{-1}\left(\frac{\beta\gamma - (1-\beta)(c+h)}{\beta(h+\gamma)}\right), \quad \hat{y}_0 = F^{-1}\left(\frac{1-\beta}{\beta\theta}\right) \tag{12.27}$$

$$y_2^* = F^{-1}\left(\frac{\beta\gamma - (1-\beta)(c+h) + \rho h}{\beta(h+\gamma)}\right), \quad \hat{y}_2 = F^{-1}\left(\frac{1-\beta-\rho}{\beta\theta}\right) \tag{12.28}$$

Note that y_1^* is the solution to:

$$\beta\gamma - (1-\beta)c - \beta(\gamma+h)F(y_1^*) - \beta\theta h F\left(\frac{hy_1^*}{r+h}\right) = 0 \tag{12.29}$$

It is easy to see that $y_0^* \leq y_2^*$ and $\hat{y}_0 \geq \hat{y}_2$. The Kuhn-Tucker condition implies (see Li et al. [2008] for details) that these parameters determine the following myopic optimal policy.

Proposition 12.5 *If there are piecewise linear default costs, the objective function is (12.22) (so the firm may borrow), and $\beta\gamma \geq (1-\beta)c$, then:*

1. *With borrowing, that is, (12.24), the solution can be determined as follows:*
   (a) *If*
   $hy_0^* - (r+h)\hat{y}_0 \geq 0$
   then $y^* = y_0^*$, $s^* = s_0^* = hy_0^* - (r+h)\hat{y}_0$, and $b^* = 0$.
   (b) *If*

$$\frac{1-\beta-\rho}{\beta\theta} \leq 1, \tag{12.30}$$

and

$$hy_0^* - (r+h)\hat{y}_0 < 0 \leq hy_2^* - (r+h)\hat{y}_2 \tag{12.31}$$

then $y^* = y_1^*$, $s^* = s_1^* = 0$, and $b^* = 0$.
   (c) *If $(1-\beta-\rho)/(\beta\theta) \leq 1$ and*

$$hy_2^* - (r+h)\hat{y}_2 < 0, \tag{12.32}$$

then $y^* = y_2^*$, $s^* = s_2^*$, and $b^* = -s^*$ where $s_2^* = hy_2 - (r+h)\hat{y}_2^*$, y_2^* and $\hat{y}_2$ are defined in (12.28).
   (d) *If*

$$\frac{1-\beta-\rho}{\beta\theta} \geq 1, \tag{12.33}$$

then $y^* = y_3^*$ where

$$y_3^* = F^{-1}\left(\frac{\beta\gamma - (1-\beta)c + \beta\theta r}{\beta(\gamma+h) + \beta\theta(r+h)}\right), \tag{12.34}$$

$s^* = -ry^*$ *and* $b^* = ry^*$ if (12.33) *holds in equality; otherwise,* $s^* = -\infty$ *and* $b^* = \infty$.

2. *Suppose that (12.30) holds, that is,* (y^*, s^*) *is* (y_0^*, s_0^*), $(y_1^*, 0)$, *or* (y_2^*, s_2^*) *then:*
   (a) y^* *increases when* γ, β, *or* ρ *increases, or when* h, c *or* θ *decreases.*
   (b) s^* *increases when* γ, h, β, θ, *or* ρ *increases, or when* c *decreases.*
   (c) b^* *increases when* c *increases or when* γ, h, β, θ, *or* ρ *decreases.*

So the optimal production/inventory policy has a base-stock structure, and this is the same as in a dynamic newsvendor model of a profit-maximizing firm without cash constraints. Each period, the production quantity is determined so that the product stock level is maintained at the level y^*. However, the optimal base-stock level for a dividend-maximizing firm with cash constraints would generally be different from the one prescribed by a standard inventory model because over-production not only implies higher holding costs or lower backorder costs but also affects the risk and the cost of default.

Moreover, the optimal cash stock level, s^*, in conjunction with y^*, balances current dividends and future retained earnings. In cases (a) and (b) of part 1 of Proposition 12.5 the marginal value of a short-term loan is negative, so the optimal cash stock level is 0. In cases (c) and (d), however, the marginal value is positive, so the target cash stock level, s^*, is positive.

Suppose that the newsvendor formulas in (12.28) determine the optimal base-stock policy (s^*, y^*). This is the case in 1(c) of the above proposition when short-term borrowing is optimal. The second formula in (12.28) sets the optimal cash stock level s^* (given the goods stock level y_2^*), or equivalently, it sets the optimal no-default limit for demand, namely the lower bound $(hy^* - s^*)/(r + h)$. If the demand is lower than the limit, then default will occur, so this might be thought of as the case of "overage" in a newsvendor model. One dollar increase in s (one dollar decrease in b) will reduce both dividend and short-term loan (a interest payment of ρ) one dollar in the current period while increasing the retained earnings by one dollar in the next period, regardless whether default occurs or not. If default occurs, then one dollar increase in s will benefit the firm by decreasing the default cost by θ (in the next period). Thus, the "overage cost" is $C_o = \beta\theta - (1 - \beta - \rho)$, the "underage cost" is $C_u = 1 - \beta - \rho$, and the optimal no-default limits for demand should be set so that the default probability equals the "critical ratio":

$$\frac{C_u}{C_u + C_o} = \frac{1 - \beta - \rho}{\beta\theta}$$

The first equation in (12.27) sets the optimal product stock level y^* in a similar fashion. Notice that a unit increase in y^* would increase the default cost by θh if the demand is lower than $(hy^* - s^*)/(r + h)$. This is the effect in addition to the usual trade-offs in a standard inventory model. Since the probability of default is already set by the above "critical ratio," this effect results in a marginal expected default cost $(1 - \beta - \rho)h$ in the first critical ratio formula regardless of demand realizations. (see Section 12.4.2 for more detailed comparisons with the standard newsvendor inventory model).

Note that if y^* is y_0^*, y_1^*, or y_2^* (cases 1(a), 1(b) and 1(c) in Proposition 12.5), then $s^* + ry^* > 0$ implies that the optimal loan is less than the value of the product base stock ($b^* = (-s^*)^+ < ry^*$), and default does not occur even when inventory is at some positive level. In this case, the sensitivity analysis results are quite intuitive. Obviously, when the loan rate ρ is high, the firm will set the cash stock level high to reduce the need to borrow and the probability of default. It is interesting to observe that when ρ is high, the firm will also set the product stock level high because a lower probability of default induces a higher product stock level. A higher γ, h, β, or θ, or a lower c implies a higher optimal cash stock level, s^*, and, hence, a smaller loan is needed to bring the total working capital to zero.

When $y^* = y_3^*$, some sensitivity analysis results do not hold. When $y^* = y_3^*$ and (12.33) holds in strict inequality, $s^* + ry^* < 0$ implies that the optimal loan is more than the value of the product base stock ($b^* = -s^* > ry^*$) and default will occur with probability one. An increase in y will have no effect on the default probability. Consequently, all results in part 2 of Proposition 12.5 remain true except one, which will change to: y^* is increasing in θ for $y^* = y_3^*$. However, this is an unlikely event in reality because lenders are unlikely to grant a loan larger than the value of the product base stock.

12.4.2 COMPARISON WITH A STANDARD NEWSVENDOR INVENTORY MODEL WITHOUT CASH CONSTRAINTS

Now we compare the operating decisions that are determined with and without financial considerations, respectively. For the operating decisions without financial considerations, we simply maximize the EPV of the profits in a standard dynamic newsvendor production/inventory model. In this model, the production decisions are made under the implicit assumption that the Modigliani-Miller Theorem is applicable to the firm. So the production decisions are made without regard to their subsequent effects on the borrowing and dividend decisions.

We make the comparison fair by letting the unit revenue r, production cost c, and inventory cost h remain the same as in the corresponding model *with* financial considerations. We assume also that the revenue net of inventory-related costs in period n is realized in period $n+1$ because period n's dividend in our model is paid before period n's revenue and inventory cost are realized.

So without financial considerations, the firm makes production decisions by maximizing the expected value of:

$$\Pi = \sum_{n=1}^{\infty} \beta^{n-1}[(r-\gamma)x_n^- - cz_n + \beta g(y_n, D_n)]$$

This is subject to the production quantity $z_n \geq 0$ for each n. Using $z_n = y_n - x_n$ and $x_n = y_{n-1} - D_{n-1}$ $(n > 1)$, this can be rewritten as:

$$\Pi = (r-\gamma)x_1^- + cx_1 + (r-\gamma-c)\sum_{n=1}^{\infty}\beta^n D_n + \sum_{n=1}^{\infty}\beta^{n-1}[-(1-\beta)cy_n - \beta(r-\gamma)y_n + \beta(r-\gamma)(y_n - D_n)^+ + \beta g(y_n, D_n)]$$

Therefore, the optimal base-stock level $\bar{y}^*$ in the standard model is a solution of the following newsvendor formula:

$$F(\bar{y}^*) = \frac{\beta\gamma - (1-\beta)c}{\beta(\gamma + h)}. \tag{12.35}$$

Incidentally, $\bar{y}^*$ can also be obtained by choosing y to maximize $K(b, s, y)$ defined by (12.10) while ignoring the default penalty ($p(\cdot)$) and financial constraints ($b_n \geq 0$; $b_n + s_n \geq 0$). Thus, we construe $\bar{y}^*$ as the production decision without considering financial costs and constraints regardless of what criterion is used. By comparing the corresponding optimality conditions, we observe that $y_0^* \leq y_1^* \leq y_2^*$ when $y^* = y_1^*$ (i.e., when y_1^* is optimal, and that $y_2^* \leq \bar{y}^*$ if $1 - \beta - \rho > 0$). The results are summarized in the following proposition.

Proposition 12.6 *Suppose that $\bar{y}^*$ is the optimal product base-stock level in the traditional inventory model determined in (12.35) and* $1 - \beta - \rho > 0$.

1. *If $y^* = y_0^*$ or y_1^* or y_2^*, then $\bar{y}^*$ is greater than or equal to y^* (i.e., $\bar{y}^* \geq y^*$).*
2. *If $y^* = y_3^*$, then $\bar{y}^* < y^*$.*

Therefore, the traditional inventory model sets the product base-stock level too high in most sensible cases (with limited borrowing). This is due to the traditional model's failure to consider the effect of the product base-stock level on the expected default caused by a mismatch between supply and demand. One unit increase in the base-stock level y has an additional effect on the expected default cost, namely, the negative effect of increasing the cash shortage due to supply surplus. In the optimality conditions, $hF[(hy^* - s^*)/(r + h)]$ measures the expected marginal cash loss in the event of a supply surplus. Because of this negative effect, the optimal base stock should be set lower than that the traditional inventory model suggests.

A unit increase in the product base-stock level y also has the positive effect of reducing whatever cash shortage would occur due to a supply shortage. When $\beta\theta \leq 1 - \beta - \rho$ (case 1(d) of Proposition 12.5), the bankruptcy penalty is small and, therefore, does not inhibit borrowing. That is, this positive effect dominates the negative one and, as a result, the optimal product base stock level is higher than in the traditional inventory model.

12.4.3 NUMERICAL EXAMPLES

The examples in this subsection illustrate the solution for a model with a one-year time period. See Li et al. (2008) for the effects of the length of the time period. Let $r = \$8$, $c = \$5$, $\gamma = \$6$, $h = \$4$ per unit per year, and $\beta = 0.8333$. So the discount rate, or opportunity cost of capital, is $1/\beta - 1 = \alpha = 0.2$. Let $\rho = 0.12$, and let the annual demand be normally distributed with mean 100 units and standard deviation 30 units. We observe that this creates an incentive to borrow because the opportunity cost of capital is higher than the interest rate on loans, i.e., $(1/\beta - 1) - \rho = 0.2 - 0.12 > 0$. Let the unit default cost θ vary from \$0.06 per dollar default to \$20 per dollar default.

Figures 12.1 and 12.2 depict the variation of the optimal product base-stock level and cash base-stock level as the unit default penalty θ increases from 0.06 to 20. Figure 12.3 shows the dependence of the global maximum of the transformed single-period payoff (the function $K(b^*, s^*, y^*)$) on the unit default penalty. If $\theta \in [0, 0.056)$ (i.e., if the default penalty is low), the optimal product base-stock level is y_3^*, which increases from 100 to 101.617 and this represents the unlikely case in 1(d) of Proposition 12.5. There, the firm borrows as much as possible. If $\theta \in [0.056, 4.472)$, the optimal product base-stock level is y_2^*, which remains at 98.315 and the optimal cash base-stock level is s_2^*, which increases from -804.937 to 0 (i.e., the loan decreases from 804.937 to 0). This is the case in 1(c) of Proposition 12.5. When θ is even greater, it is no longer optimal to borrow. If $\theta \in [4.482, 18.095)$, the optimal product base-stock level is y_1^*, which decreases from 98.315 to 93.956 and the optimal cash base-stock level remains at 0 (Proposition 12.5, 1(b)). If $\theta \in [18.00, 20)$, the optimal product base-stock level is $y_0^* = 93.943$ and the optimal cash base-stock level is s_0^*, which increases from 0 to 13.258 (Proposition 12.5, 1(a)).

In this example, the optimal product base-stock level in the standard inventory model without financial considerations is $\bar{y}^* = 100.00$. Our optimal product

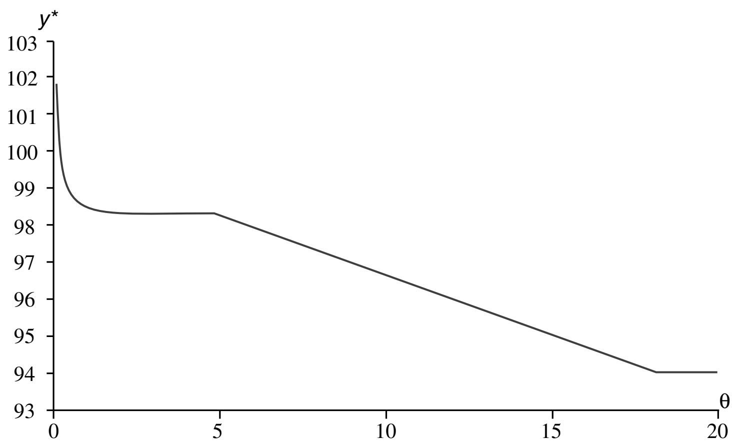


FIGURE 12.1 Optimal product base-stock level as a function of the unit default cost.

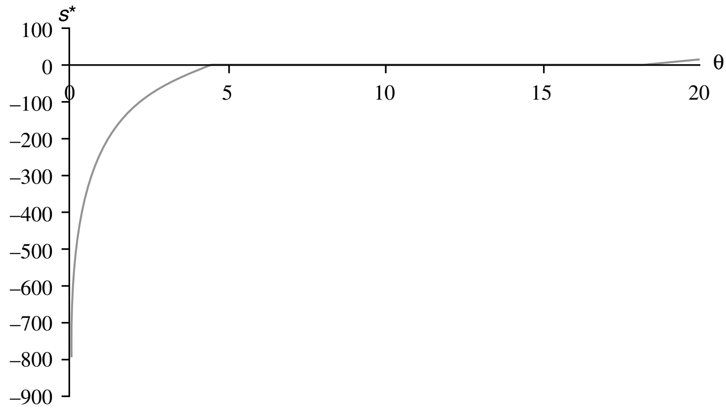


FIGURE 12.2 **Optimal cash base-stock level as a function of the unit default cost.**

base-stock level coincides with $\bar{y}^*$ only in the trivial case when there is no default penalty ($y_3^* = \bar{y}^* = 100.00$ when $\theta = 0$). In all sensible cases (with limited borrowing), $\bar{y}^*$ is always too high. In particular, it could be 6.4% higher than the optimal level, $y^* = y_0^* = 93.988$ when $\theta \geq 18.00$.

We also demonstrate here that the greatest loss of profit may be due to the lack of coordination between the levels of the product and cash stocks. It does not necessarily arise from the incorrect product base-stock level. We use the standard product base-stock level of $\bar{y}^* = 100.00$ in the preceding numerical example and compute the payoff loss in two cases. First, suppose we employ the optimal cash base-stock level s^*; that is, although the production decision is made in isolation, the firm makes the correct financial decisions. The payoff

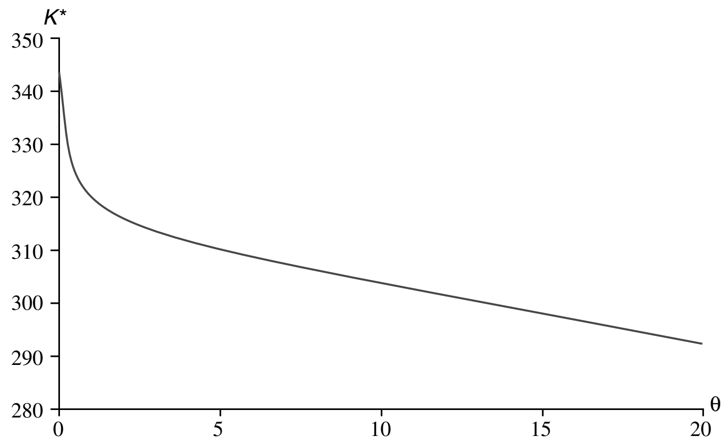


FIGURE 12.3 **Optimal transformed single-period payoff as a function of the unit default cost.**

loss, measured as a percentage of the potential contribution margin, ranges from 0.09% (when $\theta = 0.06$) to 0.96% (when $\theta = 30$). In the second case, we use a fixed cash base-stock level, for example, $s = 0$. This represents the case when the firm makes financial decisions without considering production and/or default costs. Then the payoff loss ranges from 11% at one extreme (when $\theta = 0.06$) to 3.5% at the other extreme (when $\theta = 30$).

We define the *value of coordination*, Δ, as the difference between the EPV of dividends (net of capital subscriptions) of a firm that coordinates its financial and operational decisions, and of a corresponding firm that decentralizes those decisions and makes the latter ones to maximize the EPV of profits. The *relative value of coordination* divides Δ by the EPV of dividends of the decentralizing firm.

Define K_C to be the value of the function K in (12.22) evaluated at $(s, y) = (s_2^*, y_2^*)$ plus $\beta[\beta(r - \gamma) - c]E(D)$. So $K_C/(1 - \beta)$ is the present value of the expected dividends B defined in (12.8) evaluated at the optimal coordination solution when $x_1 = w_1 = 0$. In (12.9), for simplicity of exposition, the term $\beta(\beta(r - \gamma) - c)D_n$ was separated from the terms whose conditional expectations become the function K in (12.10). Similarly, we define K_D for the value of the function K in (12.22) plus $\beta(r - \gamma - c)E(D)$, when it is evaluated at $(s, y) = (0, \bar{y}^*)$. The subscripts C and D are mnemonics for "coordinated" and "decentralized," respectively. The following parameters can be used to initiate a series of examples (in which short-term loans may be made) in which the relative values become arbitrarily large: $x_1 = w_1 = 0$, $\beta = 0.6$, $r = 10$, $c = 1$, $h = 0.5$, $\gamma = 0.8$, $\rho = 0.03$, $\theta = 1.1$, and demand is uniformly distributed on the interval $[0, 1]$.

The optimal coordinated decision with these parameters is $(s, y) = (s_2^*, y_2^*) = (-4.9, 0.52)$, the decentralized base-stock level is $\bar{y}^* = 0.1026$, $K_C = 1.54$, and $K_D = 0.54$. So the relative value is 1.85. If the other parameters remain invariant while β increases toward 0.6185, K_D drops to 0, and K_C slowly increases toward 1.62. So the relative value "explodes." That is, there are parameter settings where the decentralization EPV (the denominator) can be driven to zero while Δ (the numerator) and the coordination EPV remain strongly positive.

12.5 Long-Term Decisions on Capital Structure

In this section, we consider the firm's long-term decision on capital structure in the framework of the basic model which is developed in Section 12.2 and analyzed in Section 12.3 and Section 12.4. Assume that the capital structure of the firm has equity and long-term debt with an infinite maturity date. Denote the amount of equity by η and the amount of long-term debt by $\mu(Q)$, where Q is the amount of a periodic coupon. Assume that $\mu(\cdot)$ is concave and increasing, consistent with the capital market's perception of a rising risk of default as the coupon payment grows. This maturity-date assumption is an analytical convenience, but also it has been used in other theoretical studies and is approximated in practice. As examples, Merton (1974) uses infinite maturity debt in a dynamic model, and

Disney and IBM issued 100-year debt (cf. Leland 1994, p. 1215). Thus, in our model the firm's initial capital position is $w_1 = \eta + \mu(Q) \geq 0$.

As in the previous sections, the firm makes three decisions in each period n, the amount of the short-term loan, b_n, the quantity of physical goods to produce, z_n, and the amount of dividends to issue, v_n. Now, the interest rate for the short-term loan ρ depends on the amount of the periodic coupon payment Q and we denote this rate by $\rho(Q)$. This rate may depend too on other aspects of the model that are fixed, such as r, c, the distribution of demand (D), the amount of equity η, etc. However, there is no benefit in making such dependencies, if any, explicit in the notation. Assume that $\rho(\cdot)$ is convex and increasing with $\rho(0) \geq 0$; again, this reflects the market's perception that the risk of default rises with the size of the coupon payment. We also consider tax benefits of debt financing and let $1 - \tau$ be the marginal income tax rate.

First, we examine the optimal policies for periodic decisions regarding dividends, short-term loans, and inventory replenishment. Then we investigate the effect of the capital structure on the value of the firm.

12.5.1 OPTIMAL SHORT-TERM DECISIONS ON DIVIDEND, BORROWING, AND INVENTORY

The capital structure decision affects the dynamics of the firm's cash flow via three components: the coupon payment (Q), the interest rate for short-term loan ($\rho(Q)$), and the tax benefit (τ). With the new, additional assumptions, the flow balance equations (12.4) and (12.5) in Section 12.2 become:

$$x_{n+1} = y_n - D_n \tag{12.36}$$

$$w_{n+1} = s_n + \tau(g(y_n, D_n) - Q) \tag{12.37}$$

where

$$y_n \equiv x_n + z_n \tag{12.38}$$

$$s_n \equiv w_n - v_n + \tau[(r - \gamma)x_n^- - p(w_n) - cz_n - \rho(Q)b_n] \tag{12.39}$$

We rewrite the EPV of dividends (12.9) as:

$$\begin{aligned} E[B] &= E\left[\sum_{n=1}^{\infty} \beta^{n-1} v_n\right] \\ &= \eta + \mu(Q) + \tau(r - \gamma)x_1^- + \tau c x_1 + \tau\beta((r - \gamma - c)E[D] - Q)/(1 - \beta) \\ &\quad + \sum_{n=1}^{\infty} \beta^{n-1} K(b_n, s_n, y_n; Q) \end{aligned} \tag{12.40}$$

where

$$\begin{aligned} K(b, s, y; Q) = &-(1 - \beta)s - \tau[\beta(r - \gamma) + (1 - \beta)c]y - \tau\rho(Q)b + g(y, D) \\ &+ \tau\beta E[(r - \gamma)(y - D)^+ - p(s + \tau(g(y, D) - Q))] \end{aligned} \tag{12.41}$$

Equation (12.41) similar to $K(b, s, y)$ defined in (12.10). Thus, the dynamic program (12.13)–(12.14) becomes:

$$\psi(x; Q) = \sup_{b,s,y}\{J(b, s, y, Q) : y \geq 0, y \geq x, b \geq 0, b + s \geq 0\} \qquad (12.42)$$

$$J(b, s, y; Q) = K(b, s, y : Q) + \beta E[\psi(y - D; Q)] \qquad (12.43)$$

If the firm sells bonds with proceeds $\mu(Q)$, issues stock with proceeds η, and has an initial inventory of x_1, then the EPV of dividends is $\psi(x_1; Q) + \eta + \mu(Q) + \tau(r - \gamma)x_1^- + \tau c x_1 + \tau\beta((r - \gamma - c)E[D] - Q)/(1 - \beta)$. Since the capital structure decisions are made prior to the planning horizon that we model, adding them notationally to the model would not affect the optimality of the decisions each period regarding short-term borrowing, dividend, and production/inventory decisions. The only effect would be to add a constant to the dynamic program value function that depends on the capital structure decisions. Thus, all the results in the previous sections of this chapter remain valid in the model that is augmented with an initial capital structure. In particular, there is a myopic optimal policy in which $b_n = s_n^-$ (Propositions 12.3 and 12.4). Let $(s, y) = (s(Q), y(Q))$ maximize $K(s^-, s, y; Q)$. If $x_1 \leq y(Q)$, then the myopic policy $(b^*, s^*, y^*) = (s(Q)^-, s(Q), y(Q))$ is optimal.

To determine the optimal short-term decisions $(s(Q), y(Q))$ explicitly, we shall proceed with the special case in which the default penalty is piecewise linear, and the revenue and inventory cost functions are defined in (12.20) and (12.21).

Note that if $s + \tau(ry - Q) < 0$, then the firm will default with certainty and is unlikely to get any loans. Thus, we proceed with the interesting case $s + \tau(ry - Q) \geq 0$. With this restriction, we may rewrite the transformed single-period payoff in (12.41) as follows:

$$\begin{aligned} K(b, s, y; Q) &= -(1 - \beta)s + \tau[\beta\gamma - (1 - \beta)c]y - \tau\rho(Q)b \\ &\quad -\tau\beta(\gamma + h)\int_0^y (y - x)f(x)dx \\ &\quad -\tau^2\beta\theta\int_0^{(Q+hy-s/\tau)/(r+h)} (Q + hy - s/\tau - (r + h)x)f(x)dx. \end{aligned} \qquad (12.44)$$

Differentiating the expression gives:

$$\frac{\partial K}{\partial s} = -(1 - \beta) + \tau\beta\theta F[(Q + hy - s/\tau)/(r + h)]$$

$$\frac{\partial K}{\partial y} = \tau\left(\beta\gamma - (1 - \beta)c - \beta(\gamma + h)F(y) - \tau\beta\theta h F[(Q + hy - s/\tau)/(r + h)]\right)$$

Parallel to Proposition 12.5, we define:

$$y_0^* = F^{-1}\left(\frac{\beta\gamma - (1-\beta)(c+h)}{\beta(h+\gamma)}\right), \quad \hat{y}_0 = F^{-1}\left(\frac{1-\beta}{\tau\beta\theta}\right) \tag{12.45}$$

$$\begin{aligned} y_2^*(Q) &= F^{-1}\left(\frac{\beta\gamma - (1-\beta)(c+h) + \tau\rho(Q)h}{\beta(h+\gamma)}\right) \\ \hat{y}_2(Q) &= F^{-1}\left(\frac{1-\beta-\tau\rho(Q)}{\tau\beta\theta}\right) \end{aligned} \tag{12.46}$$

Let $y_1^*(Q)$ be the solution to equation:

$$\beta\gamma - (1-\beta)c - \beta(\gamma+h)F(y_1^*) - \tau\beta\theta h F\left(\frac{Q+hy_1^*}{r+h}\right) = 0 \tag{12.47}$$

Note that if:

$$\frac{1-\beta}{\tau\beta\theta} \leq 1 \tag{12.48}$$

then y_i^* and $\hat{y}_i$, $i = 0, 1, 2$, are well defined. Furthermore, since $F(\cdot)$ is a monotone increasing function, $y_0^* \leq y_1^*(Q) \leq y_2^*(Q)$ and $\hat{y}_0 \geq \hat{y}_2(Q)$.

Proposition 12.7 *For a given capital structure with equity η and long-term debt $\mu(Q)$, there exists a myopic short-term operating policy determined as follows.*

1. *If*

$$0 \leq Q \leq (r+h)\hat{y}_2(Q) - hy_2^*(Q) \tag{12.49}$$

then:

(a) *$y(Q) = y_2^*(Q)$, $s(Q) = s_2^*(Q)$ and $b(Q) = -s(Q)$ where $s_2^*(Q) = \tau[Q + hy_2^*(Q) - (r+h)\hat{y}_2(Q)]$ and $y_2^*(Q)$ and $\hat{y}_2(Q)$ are defined in (12.46).*

(b) *$y(Q)$ is increasing in Q and $ds(Q)/dQ \geq \tau$*

2. *If*

$$(r+h)\hat{y}_2(Q) - hy_2^*(Q) \leq Q \leq (r+h)\hat{y}_0 - hy_0^* \tag{12.50}$$

then:

(a) *$y(Q) = y_1^*$ and $s(Q) = s_1^* = 0$ where y_0^* and $\hat{y}_0$ are defined in (12.45) and y_1^* are defined in (12.47).*

(b) *$y(Q)$ is decreasing in Q.*

3. *If*

$$(r+h)\hat{y}_0 - hy_0^* \leq Q \tag{12.51}$$

then:

(a) $y(Q) = y_0^*$, $s(Q) = s_0^* = \tau[Q + hy_0^* - (r+h)\hat{y}_0]$ *and* $b(Q) = 0$ *where* y_0^* *and* $\hat{y}_0$ *are defined in (12.45).*

(b) $y(Q)$ *is constant with respect to* Q *and* $ds(Q)/dQ = \tau$.

Intuitively, since a larger long-term debt induces higher expected default costs, one would expect a more highly leveraged firm to have a lower product stock level and a higher (internally generated) cash stock level than if it had less debt. The above proposition seems to indicate that the relationship is not that simple. In particular, when long-term debt (Q) grows, the optimal product stock level rises (when $Q \le (r+h)\hat{y}_2(Q) - hy_2^*(Q)$) but then falls and remains constant thereafter. In order to understand this counterintuitive result that inventory rises with the long-term leverage, note that this occurs only when short-term borrowing is optimal and the first-order condition is as follows:

$$\tau\beta\theta F[(Q + hy - s/\tau)/(r+h)] = 1 - \beta - \tau\rho(Q), \tag{12.52}$$

$$\beta(\gamma + h)F(y) + \tau\beta\theta h F[(Q + hy - s/\tau)/(r+h)] = \beta\gamma - (1-\beta)c \tag{12.53}$$

The left side of (12.52) is the expected marginal cost (default penalty) of an additional dollar of short-term loan (or one dollar less of working capital), while the right side is the marginal benefit of funding an immediate one-dollar dividend via a short-term loan. The short-term loan should be so chosen that the marginal benefit equals the marginal cost for any given y. This yields:

$$F[(Q + hy - s/\tau)/(r+h)] = \frac{1 - \beta - \tau\rho(Q)}{\tau\beta\theta} \tag{12.54}$$

The condition sets the firm's default probability to be $[1 - \beta - \tau\rho(Q)]/(\tau\beta\theta)$, which decreases with Q and is independent of inventory decisions. Using $b(Q) = s(Q)^-$, this means that a firm with higher long-term debt would make smaller short-term loans and have a lower probability of defaulting on the short-term loans.

Substituting (12.54) into (12.53) and arranging terms, we obtain:

$$F(y) = \frac{\beta\gamma - (1-\beta)c - h(1 - \beta - \tau\rho(Q))}{\beta(\gamma + h)} \tag{12.55}$$

This is a newsvendor-like formula that determines the optimal product base-stock level. Note that Q influences the product stock level only via the optimal default probability, which is set in (12.54) by the finance decisions (short-term borrowing and dividends) independent of y, and that the probability enters in the part of the "overage cost" in the formula, $h[1 - \beta - \tau\rho(Q)]$, which is the expected marginal default cost per unit increase of y. Furthermore, this default probability is decreasing in Q and, hence, the product stock level increases in Q. So, what really happen is that a firm with more long-term debt will make smaller short-term loans, so much so that the default risk is lower and the firm is able to set a higher product stock level. This substitution of the long-term debt for the short-term loans continues until the firm depletes all its short-term debt, and then the firm must reduce the product stock level to offset the rising default risk with any further increase in long-term debt, that is, when Q satisfies the condition of Proposition 12.7 (2). If the firm has a higher long-term debt, namely, if Q satisfies the condition of Proposition 12.7 (3), then the firm would maintain a positive cash stock level and an optimal default probability, $(1-\beta)/\tau\beta\theta$, independent of Q because of no short-term borrowing. In this case, the firm again uses financial decisions (dividends) to control default risk and consequently, the optimal product stock level would be invariant with respect to Q as is the default probability.

The firm's optimal cash stock level decreases as Q increases. When short-term borrowing is optimal, the amount of the loan decreases with Q at a rate higher than τ. Then the optimal cash stock level would remain constant as Q continues to increase. Finally, when Q is sufficiently large so that the firm would keep a positive cash stock level, the firm would sacrifice a dividend of τ dollar for each increase in the long-term debt that would yield one dollar more periodic coupon payment.

12.5.2 COORDINATION VERSUS DECENTRALIZATION

We now contrast the above optimal policies when financial and operational decisions are coordinated with the optimal policies when these decisions are decentralized. There are many ways in which the decisions could be decentralized, each with a different information structure. We believe that it is sensible to consider the following definition of decentralization: (i) operational decisions are made without financial considerations as in Section 12.4.2 while (ii) financial decisions such as short-term borrowing and dividends are chosen to maximize the EPV of dividends, *treating as exogenous the cash flows that result from the operational decisions derived in (i).* This represents a situation of "one-sided coordination": operational decisions are made independently and financial decisions optimize the value of the firm to the stockholders given the outcomes of the operational decisions.

We have shown in Section 12.4.2 that the decentralized operations decision should be a base-stock policy determined in (12.35), that is, with a product base-stock level of:

$$\bar{y}^* = F^{-1}\left(\frac{\beta\gamma - (1-\beta)c}{\beta(\gamma+h)}\right)$$

We shall use a superscript "D" for the decentralized decisions in the following proposition.

Proposition 12.8 *For a given capital structure with equity η and long-term debt $\mu(Q)$, the decentralized decisions are: $y^D = \bar{y}^*$ and:*

1. *If $0 \le Q \le (r+h)\hat{y}_2(Q) - h\bar{y}^*$, then $s^D(Q) = \tau[Q + h\bar{y}^* - (r+h)\hat{y}_2(Q)]$ and $b^D(Q) = -s^D(Q)$ where $\hat{y}_2(Q)$ are defined in (12.46).*

2. *If* $(r+h)\hat{y}_2(Q) - h\bar{y}^* \le Q \le (r+h)\hat{y}_0 - h\bar{y}^*$, *then* $s^D(Q) = b^D(Q) = 0$ *where* $\hat{y}_0$ *are defined in (12.45).*
3. *If* $(r+h)\hat{y}_0 - h\bar{y}^* \le Q$, *then* $s^D(Q) = \tau[Q + h\bar{y}^* - (r+h)\hat{y}_0]$ *and* $b^D(Q) = 0$.

Proposition 12.6 shows that $\bar{y}^* \ge y_i^*$ for $i = 0, 1, 2$. As a direct corollary, we have:

Corollary 12.2 *For all Q:*

1. $y(Q) \le y^D(Q)$.
2. $s(Q) \le s^D(Q)$ *and* $b(Q) \ge b^D(Q)$.
3. *The default probability is smaller with coordinated decisions than with decentralized ones.*

This means that coordinated decisions, in comparison to decentralized decisions, entail less product stock, less (internally generated) cash stock, larger short-term loans, and lower default risk. Of course, without considering the costs of coordination (and/or the benefits of decentralization), coordinated decisions would also lead to higher expected dividend net of capital subscription. The causes of the differences are as follows. As noted in Section 12.4.2, a decentralized firm fails to consider the negative effect of the product stock level on the expected default cost and, hence, sets the product stock level too high. As a result, the firm would set the cash stock level higher than necessary and thus make short-term loans that are too low. Furthermore, without explicitly recognizing the financial conditions, particularly the long-term debt level, the firm keeps the goods level $\bar{y}^*$ invariant with respect to Q. Holding excessive illiquid goods, the firm would default in circumstances where a firm that coordinates its decisions would remain solvent; so the firm with decentralized decisions also has higher default risk.

12.5.3 OPTIMAL CAPITAL STRUCTURE

Let $B(Q)$ denote the present value of the dividends if the initial capital structure entails a periodic coupon payment Q when the firm employs the optimal short-term policies $(y_n, s_n) = (y(Q), s(Q))$ for all n and let $\mathcal{B}(Q) = E[B(Q)]$. Similarly, $B_D(Q)$ denotes the present value of the dividends if the initial capital structure entails a periodic coupon payment Q and the short-term decisions are decentralized and let $\mathcal{B}_D(Q) = E[B_D(Q)]$. It can be shown that the domain of $\mathcal{B}_D(\cdot)$ is a convex set included in the domain of $\mathcal{B}(\cdot)$, which is also a convex set. Furthermore, both $\mathcal{B}(\cdot)$ and $\mathcal{B}_D(\cdot)$ are concave in their respective domains. Let Q^* and Q^D maximize $\mathcal{B}(\cdot)$ and $\mathcal{B}_D(\cdot)$ respectively. It is straightforward to show that $\mathcal{B}(Q) \ge \mathcal{B}_D(Q)$ for each Q and $\mathcal{B}(Q^*) \ge \mathcal{B}_D(Q^D)$ since $\mathcal{B}(Q)$ is the maximal value of an optimization problem in which $(y, s) = (y^D(Q), s^D(Q))$ is feasible for each Q. That is, the coordinated short-term decisions yield a higher

EPV of firm value than the decentralized ones. We can also show that the coordinated short-term decisions require less capital under mild conditions (i.e., $Q^* \le Q^D$).

Proposition 12.9 *If β is sufficiently large and*

$$\mu'(Q)[1 - \tau\rho(Q)] < 1, \tag{12.56}$$

then

1. $Q^* \le Q^D$,
2. $s(Q^*) \le s^D(Q^D) = \tau[Q^D + h\bar{y}^* - (r+h)\hat{y}_2(Q^D)] \le 0$ *and* $b(Q^*) \ge b^D(Q^D) = -s^D(Q^D) \ge 0$ *where $\hat{y}_2(Q)$ are defined in (12.46)*

Proposition 12.9 asserts that the firm with coordinated short-term decisions has lower optimal long-term debt while its short-term loans are higher. Furthermore, the optimal leverage Q^* (or Q^D) is always set at a level that is sufficiently low so that the short-term borrowing is optimal regardless whether short-term decisions are coordinated or decentralized.

To appreciate that condition (12.56) is not restrictive, consider how much additional capital the firm could raise via long-term debt if the coupon payment were $Q + dQ$ instead of Q. Since there is a possibility of default, a prudent investor would provide an increment that is at most the present value of the time stream of riskless coupon increments. That is, $\mu(Q + dQ) - \mu(Q) \le dQ[1 - \tau\rho(Q)]^{-1}$. This is reduced to (12.56) when $dQ \to 0$.

The Modigliani-Miller Theorem states that a firm's capital structure does not affect its market value and therefore, the operations decisions can be separated from finance decisions if the capital market is perfect and complete with equal borrowing and lending rates and in the absence of taxes, bankruptcy costs and asymmetric information. Our model fails to satisfy these conditions because it incorporates taxes, default costs, and a discrepancy between the firms borrowing and lending interest rates. Since we normalize the lending rate to zero, the optimal coordinated decisions should converge to the decentralized decisions as the default penalty and taxes fade away and the borrowing rate goes to zero. Let $\theta \to 0$, $\tau \to 1$ and $\rho(\cdot) = 0$. Then the expected single-period payoff (12.44) becomes:

$$K(b, s, y) = -(1-\beta)s + (\beta\gamma - (1-\beta)c)y - \beta(\gamma + h)\int_0^y (y - x)f(x)dx$$

This is independent of Q. Optimizing the above expression with respect to y yields $y(Q) = y^D(Q) = \bar{y}^*$. This is the same inventory level as in Section 12.4.2 where short-term decisions are decentralized. The Modigliani-Miller world has been restored.

12.6 Extensions and Variations of the Basic Model

Many possible extensions and/or variations can be derived from our basic model developed in Section 12.2. We discuss only two of them in this section. We first show how our basic model can be modified to incorporate the "wipeout" bankruptcy, which results in termination of the firm. The second extension deals with a situation in which capital subscriptions are precluded (i.e., with the additional constraint that dividends must be non-negative). We show that the key results from the basic model remain valid (qualitatively) except the optimal policy with non-negative dividends is no longer myopic.

12.6.1 WIPEOUT BANKRUPTCY

The notion of bankruptcy in the basic model is consistent with Chapter 11 of the U.S. Bankruptcy Code and generates substantial costs of reorganizing the firm. In this subsection we consider an alternative version of bankruptcy that is consistent with Chapter 7, namely dissolution of the firm. We arrive at three important insights by maximizing the EPV of dividends prior to dissolution. First, the key features of an optimal policy with "reorganization" bankruptcy remain valid with "wipeout" bankruptcy. These are the myopic optimum property and Propositions 12.1, 12.2, 12.3, and 12.4 (with minor changes). Second, the firm should be more short-sighted because its discount factor is reduced from β to $\beta q(s, y)$ (recall that $q(s, y)$ is the probability that bankruptcy does *not* occur in period $n+1$ if $s_n = s$ and $y_n = y$). In effect, the firm's choices influence its time preference as well as conversely. Third, the life-time of the optimally operated firm has a geometric probability distribution.

Let T be the lifetime of the firm, so $T = \sup\{n : w_n > 0\}$, and we maximize $E[B]$ where $B = \sum_{n=1}^{T} \beta^{n-1} v_n$. The substitutions that lead from (12.8) to (12.9) yield:

$$\begin{aligned} B = \sum_{n=1}^{T-1} \beta^{n-1} [&\beta(r-\gamma)(y_n - D_n)^+ + \beta g(y_n, D_n) + \beta(r-\gamma-c)D_n \\ &-(1-\beta)(s_n + cy_n) - \rho b_n] + (r-\gamma)x_1^- + cx_1 \\ &+w_1 - \beta^T(s_T + cy_T + \rho b_T) \end{aligned} \tag{12.57}$$

Therefore, an optimal coordinated policy for making operating and financial decisions maximizes $E(B_0)$ where:

$$\begin{aligned} E(B_0) = &E[\sum_{n=1}^{T-1} \beta^{n-1}[\beta(r-\gamma)(y_n - D_n)^+ + \beta g(y_n, D_n) \\ &+ \beta(r-\gamma-c)D_n - (1-\beta)(s_n + cy_n) - \rho b_n] \\ &- \beta^{T-1}(s_T + cy_T + \rho b_T) \end{aligned} \tag{12.58}$$

This model is a generalization of an inventory process with a stopping time, so the results in Lovejoy (1992) bound the error that would result from using the policy identified in Section 12.3 rather than a policy that optimizes (12.58). However, an approximation is unnecessary because the model with wipeout bankruptcy satisfies the condition in Sobel (1981) and, therefore, has an optimal myopic solution.

Earlier in Section 12.4 we introduced the notation $\Gamma(s, y)$ for the set of outcomes of demand D_n which do *not* precipitate bankruptcy in period $n+1$ if $s_n = s$ and $y_n = y$; namely $\Gamma(s, y) = \{d : s + g(y, d) > 0\}$. Also, recall the notation $q(s, y)$ for the probability that bankruptcy does *not* occur in period $n+1$ if $s_n = s$ and $y_n = y$, that is, $q(s, y) = P\{D \in \Gamma(s, y)\}$. Then:

$$\begin{aligned} E(B_0) &= \sum_{n=1}^{\infty} \beta^{n-1}\Big(- P\{T = n\}E[s_T + cy_T + \rho b_T | T = n] \\ &\quad + P\{T > n\}E[\beta(r-\gamma)(y_n - D_n)^+ + \beta g(y_n, D_n) \\ &\quad + \beta(r - \gamma - c)D_n - (1-\beta)(s_n + cy_n) - \rho b_n | T > n]\Big) \\ &= \sum_{n=1}^{\infty} \beta^{n-1} \prod_{k=1}^{n-1} q(s_k, y_k) K_0(b_n, s_n, y_n) \end{aligned} \tag{12.59}$$

where

$$\begin{aligned} K_0(b, s, y) = q(s, y)E[&\beta(r-\gamma)(y-D)^+ + \beta g(y, D) + \beta(r-\gamma-c)D \\ &-(1-\beta)(s + cy) - \rho b | D \in \Gamma(s, y)] \\ &-(s + cy + \rho b)(1 - q(s, y)) \end{aligned} \tag{12.60}$$

The objective was (12.11) when bankruptcy signified reorganization. The only difference between (12.11) and (12.59) is that the single-period discount factor increases from $\beta q(s, y)$ to β. Therefore, dynamic program (12.13) remains valid when (12.14) is replaced with:

$$J(b, s, y) = K_0(b, s, y) + \beta q(s, y)E[\psi(y - D) | D \in \Gamma(s, y)]. \tag{12.61}$$

Thus, *the myopic optimum property and Propositions 12.2, 12.3, and 12.4 remain valid (with minor changes) when bankruptcy signifies dissolution of the firm.*

The preceding observation leads to a geometric probability distribution for the lifetime of the firm; *this is a testable hypothesis*. Let (s^*, y^*) globally maximize $K_0((-s)^+, s, y)$ subject to $y \geq 0$. If $x_1 \leq y^*$, it is optimal for $(b_n, s_n, y_n) = ((-s^*)^+, s^*, y^*)$, $n = 1, \ldots, T$. Therefore:

$$P\{T = n\} = [1 - q(s^*, y^*)]q(s^*, y^*)^{n-1}$$

Proposition 12.10 *If $x_1 \leq y^*$, it is optimal for $(b_n, s_n, y_n) = ((-s^*)^+, s^*, y^*)$, $n = 1, \ldots, T$, and, as a consequence, T has a geometric distribution with parameter $q(s^*, y^*)$.*

12.6.2 NON-NEGATIVE DIVIDENDS

Large publicly traded firms cannot ordinarily obtain capital subscriptions because they have limited liability stockholders. So here we briefly analyze the model without capital subscriptions (i.e., with the constraint that dividends must be non-negative). If we add the constraint $v_n \geq 0$ to the formulation of the basic model in Section 12.2, the objective remains (12.11) and because $v_n = w_n + (r - \gamma)x_n^- - p(w_n) - s_n - cy_n + cx_n - \rho b_n$, the constraints (12.12) are augmented with:

$$s_n + cy_n + \rho b_n \leq w_n + (r - \gamma)x_n^- - p(w_n) + cx_n$$

This necessity causes an additional state variable in the dynamic program. That is, the dynamic program for the model in Section 12.2 has a scalar state, the inventory level x, that must be augmented now with w, the amount of retained earnings at the beginning of a period, because this new constraint depends on both state variables. The resulting dynamic program, instead of (12.15) and (12.16), is the following recursion with $\psi_{N+1}(\cdot, \cdot) \equiv 0$ for each $n = 1, 2, \ldots, N$, $x \in \Re$, and $w \in \Re$:

$$\psi_n(w, x) = \max_{b,s,y}\{J_n(b, s, y) : y \geq (x)^+,\, b \geq 0,\, b + s \geq 0,$$
$$s + cy + \rho b \leq w + (r - \gamma)x^- - p(w) + cx\} \qquad (12.62)$$
$$J_n(b, s, y) = K(b, s, y) + \beta E\big(\psi_{n+1}[s + g(y, D), y - D]\big) \qquad (12.63)$$

Let $b_n(w, x)$, $s_n(w, x)$, and $y_n(w, x)$ be the optimal values of b, s, and y, respectively in (12.62). We show that even with the non-negativity of dividends in force, the results analogous to Proposition 12.2 and 12.4 still hold.

Proposition 12.11 *If $p(\cdot)$ is a decreasing convex function on $\Re$ and $g(\cdot, d)$ is a concave function on $\Re$ for each $d \geq 0$, then:*

1. *The value function in (12.62), $\psi_n(\cdot, \cdot)$, is a concave function on $\Re^2$ and $J_n(\cdot, \cdot, \cdot)$ in (12.63) is a concave function on $\Re^3$ for each n.*
2. *For each n, $y_n(w, x)$, $z_n(w, x) = x - y_n(w, x)$, $v_n(w, x)$ and $s_n(w, x)$ are increasing with respect to $w \in \Re$ and $x \in \Re$. So $b_n(w, x)$ is a decreasing function of w and x.*
3. *$b_n(w, x) = (-s_n(w, x))^+$ is optimal for all $n = 1, 2, \ldots$ and $(w, x) \in \Re^2$.*

The following result compares the optimal policy when capital subscriptions are allowable with the optimal policy when dividends are constrained to be non-negative. Recall the notation $b_n(x)$, $s_n(x)$, and $y_n(x)$ for the optimal amounts of the short-term loan, residual retained earnings, and product base-stock level, respectively, when the inventory level is x and n periods remain in the planning horizon.

Proposition 12.12 *Under the assumptions of Proposition 12.11, for each n, x, and w,*

- $y_n(x, w) \leq y_n(x)$ *and* $s_n(x, w) \leq s_n(x)$ *(so* $b_n(x, w) \geq b_n(x)$*)*
- *As w grows,* $y_n(w, x) \to y_n(x)$ *and* $s_n(w, x) \to s_n(x)$ *(so* $b_n(w, x) \to b_n(x)$*)*

That is, a firm that optimally coordinates its operational and financial decisions but cannot mandate capital subscriptions has lower inventories and higher short-term loans than its counterpart, which may obtain capital subscriptions if it wishes. Therefore, each period the former firm has a higher probability of default than the latter. The latter firm can turn to a capital subscription *or* a short-term loan, whereas the former firm can increase liquidity only with the loan, so its level of residual retained earnings is lower and its short-term loan is higher. Similarly, without recourse to capital subscriptions, the former firm has a lower product base-stock level because it is less prone to buy or produce goods.

12.7 Concluding Remarks

In this chapter, we first formulate and analyze a dynamic stochastic model of the firm that coordinates its operational and financial decisions with the criterion of the EPV (expected present value) of the time stream of dividends received by a representative share owner. The model is surprisingly tractable because it has a myopic policy that is optimal and, therefore, susceptible to analysis. So we can contrast the result with inventory policies based on models without financial considerations. It turns out that the profit-maximizing product base-stock level is always higher than the dividend-maximizing base-stock level because the former fails to consider the effect of the product base-stock level on the default risk. Numerical examples show that the opportunity cost of detaching the two functions, measured in dividends, can be significant. In particular, we find that this cost of decentralization can sometimes be due primarily to the uncoordinated financial decisions rather than erroneous inventory base-stock levels. Since our basic model considers Chapter 11 bankruptcy and admits capital subscriptions, we also show that our analysis and results can be extended to incorporating Chapter 7 bankruptcy and precluding capital subscriptions.

We then investigate the relationship between capital structure and short-term financial and operations decisions. We assume that the firm's capital structure is determined by a decision on equity and long-term debt that is made at the beginning and imposes constraints on the short-term decisions such as production, dividends, and short-term borrowing as in our basic dynamic stochastic model. Long-term debt and short-term loans are tax sheltered but entail a bankruptcy risk and hence, affect operations. We consider both the coordination of all short-term decisions and a particular "decentralization" case in which the firm employs the standard profit-maximizing inventory policy without financial considerations and coordinates other short-term decisions. At the short-term operating level, the profit-maximizing base-stock policy causes a firm to have too much inventory, to carry excess working capital, to make smaller short-term loans and, consequently, to issue smaller dividends.

We show that a firm with higher long-term debt will reduce its short-term loan to the extent that the default risk is lower and the firm is able to set a higher product stock level. This substitution of the long-term debt with the short-term loans continues until the firm depletes all its short-term debt, and then the firm must reduce its product stock to offset the rising default risk with any further increase in long-term debt. Thus, as the amount of long-term debt increases parametrically, the optimal product stock level increases initially but then decreases. On the other hand, the short-term loan decreases and the cash stock level increases for either coordinated or decentralized cases except the rates of changes are smaller when the short-term decisions are decentralized.

A particular capital structure affects short-term decisions and, therefore, dividends and the market value of the firm. Anticipating these effects, we characterize the optimal amount of leverage for the firm. The optimal amount of long-term debt is smaller if the firm coordinates its short-term decisions than if it decentralizes them. Furthermore, the optimal amount of long-term debt will never be so high that the short-term loan is precluded.

In this chapter the results were obtained with a finite-horizon model, but they are inherited by the analogous infinite-horizon model. As $N \to \infty$, under reasonable assumptions, the functions in (12.15) and (12.16) (and in [12.62] and [12.63]) converge point-wise to well-behaved limits which satisfy the obvious infinite-horizon analog of (12.15) and (12.16) (and (12.62) and (12.63)). If $x_1 \leq y^*$, it follows from Proposition 12.5 that the pairs of dividends and production quantities in successive periods (starting with the second period) are independent and identically distributed random vectors. However, within a period, the dividend and production quantity are not independent.

A natural extension of this work is to employ pricing decisions of the firm as an important way of correcting inventory and cash flow problems. We leave this complication to a future investigation. More generally, the modern firm is a complex multiproduct institution and, even with modern computational methods, quantitative strategic analysis is crude at best. We believe that there is still great potential for considerable value-added investigation internal to the firm and the model investigated here exemplifies this observation.

12.8 Bibliographical Notes

The literature on models of production and inventory systems (cf. Graves et al. 1993) is large but, until recently, financial considerations were conspicuously absent from most models of operational decisions. An early analysis of financial considerations (Shubik and Thompson 1959) is based on a controlled random walk model of the dynamics of the growth of a dividend-paying non-financial firm between a reflecting and an absorbing barrier. The reflecting barrier is created by dividend payments and the absorbing barrier by bankruptcy conditions which put the firm out of business. One way of reflecting the different goals of managers and owners is to separate ownership and management in a game of economic survival. This clarifies the roles of dividends and bankruptcy. This chapter's model embodies strategic aspects of the trade-offs between bankruptcy and paying dividends. Alternatively, if the firm is penalized by an insolvency, but is in a position to continue to operate, this too can be modeled as a reflecting barrier. That is, there may not be as great a difference between the effects of reorganization and wipeout bankruptcy as many have supposed. See Radner and Shepp (1996) for recent work of this kind. Hadley and Whitin (1963) and Sherbrooke (1968) are early papers on inventory management with budgetary constraints. Another research connection between operations and finance is the treatment of the maintenance of cash safety levels as an inventory control problem (cf. Porteus 1972, Shubik and Sobel 1992).

The synthesis of operational and financial considerations is primarily recent in origin. Some papers model operational decisions in the presence of foreign exchange exposure (e.g., Kogut and Kulatilaka 1994, Huchzermeier and Cohen 1996, Dasu and Li 1997, Aytekin and Birge 2004, Dong et al. 2006). Other papers use approaches similar to ours to analyze capacity-expansion problems with financial constraints (e.g., Birge 2000, Van Mieghem 2003, Babich and Sobel 2004).

Some recent research on the coordination of operational and financial decisions can be partitioned according to whether or not the models and criteria are influenced by the capital structure of the firm. Perhaps the earliest papers without capital structure considerations are Archibald et al. (2002), which optimizes the probability of survival of a growing firm that manages an inventory, and Buzacott and Zhang (2004) who model a growing manufacturer that finances production both with loans secured by inventory and with unsecured loans. The latter paper demonstrates the importance of jointly considering production and financing decisions in a dynamic deterministic setting and proposes a single-period newsvendor model to examine the incentives for a lender and a borrower to engage in asset-based financing.

Capital structure is conspicuous in the static model that is essentially common to Xu and Birge (2004), Xu and Birge (2005), and Dada and Hu (2008). There, a financially constrained firm coordinates its operating decisions and short-term borrowing. Xu and Birge (2004), assuming that the creditor is nonstrategic, note that bankruptcy costs remove the firm from the Modigliani-Miller world. So the firm's borrowing capacity is limited, and it may have to produce less than the unconstrained optimum. Comparative statics implies that the firm's optimal production quantity decreases as its debt increases. The authors conclude that "a low-margin company should select a conservative output level and an aggressive financial decision, while a high-margin company" should do the opposite.

Xu and Birge (2005) numerically examine how the value of the firm depends on production unit cost and debt-equity ratio, and then compare their results with market data. They find that the model predicts lower debt-equity ratios than in the data, and they attribute this difference to the absence of long-term debt in the model. They show that a low-margin producer faces higher agency costs than a high-margin one. Dada and Hu (2008) assume that the creditor *is* strategic and use game theory to show (a) that the firm produces less than the unconstrained optimum even without bankruptcy costs, and (b) the optimal production quantity increases as a function of the firm's equity.

The model that underlies these three papers is static and yields conclusions that differ from some of those in Hu and Sobel (2008), which is based on a dynamic model that distinguishes between long-term debt and short-term loans. In that paper, as the *long-term* debt increases, (a) the optimal product stock level increases and then decreases, and (b) the optimal short-term loan decreases. The optimal production quantity (in any period after the first) is the previous period's demand, and this is typically true of dynamic newsvendor-like models without capacity constraints. So the optimal product stock levels are driven by the amount of long-term debt, but production amounts do not depend on it at all.

The firm in the valuation model in Xu and Birge (2006) maximizes a combination of the EPV of the net cash flow to shareholders and multiples of other firm attributes, and it decides whether to continue to produce or to default and liquidate the firm. The contingent default opportunity raises the EPV of the net cash flow to shareholders, so it yields higher equity valuations than traditional valuation and planning models. The authors conclude that valuations are understated if they stem from models which exclude the possibility of contingent defaults.

We do not examine the agency issues of nonowner-managers, but see Xu and Birge (2005) for their inclusion. Real owner-managers have shorter time horizons than our "firm;" Babich and Sobel (2004) optimize a cash-out goal of an owner-manager.

Other papers leverage the results in this chapter but change the assumptions in Section 12.2. Brunet and Babich (2007) compute the signaling value of trade credit financing for the acquisition of goods.

Nonlinear production costs in Sobel and Zhang (2003) lead to more complicated optimal inventory policies than the base-stock levels in this chapter. Sobel and Turcic (2007) consider how best to adapt to evolving market conditions that are modeled with a more flexible and realistic model of demand than in Section 12.2. That raises issues that cannot be addressed here, such as the effects of firm growth on the optimal levels of product and cash stock.

REFERENCES

Archibald T.W., Thomas L.C., Betts J.M., Johnston R.B. 2002. Should start-up companies be cautious? Inventory policies which maximize survival probabilities. *Management Science*, **48** (9):1161–1174.

Aytekin U., Birge J.R. 2004. Optimal investment strategies and production across markets with stochastic interest rates. Working Paper.Evanston, Illinois: Department of Industrial Engineering and Management Sciences, Northwestern University.

Babich V., Sobel M.J. 2004. Pre-IPO Operational and Financial Decisions. *Management Science*, **50** (7):935–948.

Birge J.R. 2000. Option methods for incorporating risk into linear capacity planning models. *Manufacturing and Service Operations Management*, **2**:19–31.

Brunet P.-Y., Babich V. 2007. Inventory of cash-constrained firms and the option to acquire future financing. Working Paper. Ann Arbor, Michigan: Department of Industrial and Operations Engineering, University of Michigan.

Buzacott J., Zhang R. 2004. Inventory management with asset-based financing. *Management Science*, **50** (9):1274–1292.

Dada M., Hu Q. 2008. Financing newsvendor inventory. *Operations Research Letters*, **36**:569–573.

Dasu S., Li L. 1997. Optimal operating policies in the presence of exchange rate variability. *Management Science*, **43**:705–722.

Dong L., Kouvelis P., Su P. 2006. Operational hedging strategies and competitive exposure to exchange rates. Working Paper. St. Louis, Missouri: Olin School of Business, Washington University.

Graves S.C., Rinnoy Kan, A.H.G., Zipkin, P.H., editors. 1993. *Handbooks in Operations Research and Management Science, Vol.4 of Logistics of Production and Inventory*. Amsterdam, The Netherlands: North-Holland.

Hadley G.J., Whitin T.M. 1963. *Analysis of Inventory Systems*, Englewood Cliffs, New Jersey: Prentice Hall.

Hu Q.(J.), Sobel M.J. 2005. Capital structure and inventory management. Unpublished working paper. Cleveland, Ohio: Weatherhead School, Case Western Reserve University.

Huchzermeier A., Cohen M.A. 1996. Valuing operational exibility under exchange rate uncertainty. *Operations Research*, **44**:100–113.

Kogut B., Kulatilaka N. 1994. Operational flexibility, global manufacturing, and the option value of a multinational network. *Management Science*, **40** (1):123–129.

Leland H.E. 1994. Corporate debt value, bond covenants, and optimal capital structure. *The Journal of Finance*,**49**:1213–1252.

Li L., Shubik M., Sobel M.J. 2008. Production with dividends and default penalties. Unpublished working paper series B. New Haven, Connecticut: Operations Research, Yale School of Management. June 28, 1997;revised December 28, 2008.

Mendelssohn R., Sobel M.J. 1980. Capital accumulation and the optimization of renewable resources. *Journal of Economic Theory*, **23** 243–260.

Merton R.C. 1974. On the pricing of corporate debt: The risk structure of interest rates. *The Journal of Finance*, **29**:449–470.

Modigliani F., Miller M.H. 1958. The cost of capital, corporation finance, and the theory of investment. *American Economic Review*, **48**:261–297.

Porteus E.L. 1972. Equivalent formulations of the stochastic cash balance problem. *Management Science*, **19**:250–253.

Radner R., Shepp L. 1996. Risk vs. profit potential: A model for corporate strategy. *Journal of Economic Dynamics and Control*, **20**:1373–1393.

Sherbrooke C.C. 1968. Metric: Multiechelon technique for recoverable item control. *Operations Research*, **16**:122–141.

Shubik M., Sobel M.J. 1992. On matching book: A problem in banking and corporate finance. *Management Science*, **38**:827–839.

Shubik M., Thompson G.L. 1959. Games of economic survival. *Naval Research Logistics Quarterly*, **6**:111–123.

Sobel M.J. 1981. Myopic solutions of Markov decision processes and stochastic games. *Operations Research*, **16**:995–1009.

Sobel M.J., Turcic D. 2007. Coordination of inventory, capital, and dividends in nascent firms. Unpublished working paper. Cleveland, Ohio: Weatherhead School, Case Western Reserve University.

Sobel M.J., Zhang J. 2003. Inventory replenishment with a financial criterion. Unpublished manuscript.

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

Xu X., Birge J.R. 2004. Joint production and financing decisions: Modeling and analysis. Working paper. Evanston, Illinois: Northwestern University.

Xu X., Birge J.R. 2005. Operational decisions, capital structure and managerial compensation: A newsvendor perspective. Working paper. Evanston, Illinois: Northwestern University.

Xu X., Birge J.R. 2006. Equity valuation, production, and financial planning: A stochastic programming approach. *Naval Research Logistics Quarterly*, **53**:641–655.

CHAPTER THIRTEEN

Bank Financing of Newsvendor Inventory: Coordinating Loan Schedules

QIAOHAI (JOICE) HU AND MAQBOOL DADA

We examine how a capital-constrained newsvendor's procurement and external financing may be affected by its own internal capital level. We focus on a Stackelberg game in which the newsvendor is the follower and the bank is the leader. This setting captures instances when specialized knowledge precludes market efficiency. Our results show that if the cost of borrowing is not too high, the capital constrained newsvendor borrows funds to procure an amount that is somewhat less than would be ideal. In return the bank charges an interest rate that decreases in the equity position of the firm. We also observe that holding the mean fixed, as demand becomes less skewed, the financial channel's profits tend to rise, and the equilibrium leads to relatively less loss in channel efficiency. Strategies for achieving the first-best solution are also developed.

Acknowledgement: This treatise is an extended and updated version of "Financing Newsvendor Inventory" published in *Operations Research Letters, 2008, vol.36, 569-573.*

Handbook of Integrated Risk Management in Global Supply Chains, First Edition. Edited by Panos Kouvelis, Lingxiu Dong, Onur Boyabatli, and Rong Li.

13.1 Introduction

When faced with uncertain demand, effective inventory management entails procuring a sufficient amount of stock to buffer against variations in demand from its mean. Typically, the cost of holding inventory, including its financing, is balanced against other costs that include those of order processing and order fulfillment. In the vast literature on this fundamental topic, researchers almost inevitably assume that while inventory is costly, there is sufficient working capital available for its financing. Such a prototypical case arises when working capital is provided by corporate headquarters of large firms to finance divisional inventory at a prespecified hurdle rate.

While a successful large corporation may be in a position to adequately fund inventory, there are many situations in which decision makers face liquidity constraints. An important case arises when a firm is nascent and is therefore, necessarily, liquidity-constrained. The recent documented difficulties of PenAgain, one such start-up, are detailed in a story by Bounds (2006.) in *The Wall Street Journal*. For complementary reasons, a firm under financial distress or bankruptcy protection also finds itself capital-constrained. In such cases the firm resorts to factors or debt-in-place financing to fund its illiquid operations. And, in many international transactions, otherwise well-financed entities find it convenient to purchase goods by executing letters-of-credit, which bind financial intermediaries to ensure that suppliers honor purchase agreements.

The gist of these situations is that the procurement of inventory is made by a mixture of equity and debt. Some natural questions that arise include: (1) How does the bank determine the interest rate?, (2) How much should be borrowed?, (3) How is this transaction influenced by the equity available?, and, (4) How are the terms of the transaction influenced by the nature of demand uncertainty? The goal of this paper is to study the structure of such a transaction by focusing on the case of a one-time purchase of a seasonal good.

We use the setting of the inventory procurement problem of a capital-constrained newsvendor (CCNV). Given a set of purchase and selling prices, and knowing the distribution of seasonal demand, the CCNV is unable to order a quantity that satisfies the newsvendor fractile. Therefore, it must decide how much to borrow at a given interest rate to finance additional procurement. Since demand is uncertain, it is possible that sales will not be high enough to payoff the loan at the end of the selling season. Thus, the bank must take into account the possibility of default when determining the interest rate to charge.

We represent the strategic interaction between the two parties as a Stackelberg game in which the bank is the leader and the CCNV is the follower. This setting captures instances when specialized knowledge precludes existence of market efficiency. Our results show that if the cost of borrowing is not too high, the CCNV borrows funds to procure an amount that is less than the channel optimal quantity. In return the bank charges an interest rate that decreases in the equity position of the firm. Our numerical work further suggests that efficiency is relatively high and tends to increase with decreasing skewness of the demand distribution, with the mean fixed. However, this loss in efficiency can be eliminated if a profit-splitting nonlinear loan mechanism is invoked.

Until recently financial considerations were conspicuously scarce in the extensive literature on models of inventory and production processes. However, research on the coordination of operational and financial decisions has been growing. Start-up firms with a criterion of maximizing survival probability, according to Archibald et al. (2002), should be more cautious in their component purchasing strategy than well-established ones, and their purchasing quantity is not necessarily monotone in their available capital. Babich and Sobel (2004) examine the capacity expansion decisions of a firm so that its expected proceedings from an initial public offering of stock is maximized. Li et al. (1997) study the coordination of operations with finance in a dynamic setting. Birge (2000) works on the impact of risk on capacity planning models, Xu and Birge (2006) incorporate bankruptcy deadweight costs in the context of multiperiod inventory models, and Kogut and Kulatilaka (1994), Huchzermeier and Cohen (1996), and Kouvelis and Li (1999) study the value of operational flexibility, and Gaur and Seshadri (2005) study hedging demand risk in inventory decisions. Lederer and Singhal (1994) and the more recent work, Boyabat and Toktay (2010), examine the joint financing and capacity investment problem. More recently, Gupta and Wang (2009) study a dynamic inventory model in which the retailer's inventory is financed with trade credits offered by the supplier. The authors show that the base-stock policy continues to be optimal. Our work is also related to the supply chain coordination literature on which Cachon (2005) is an excellent review.

The most closely related papers are Xu and Birge (2004), Buzacott and Zhang (2004). In a nonstrategic framework, Xu and Birge (2004) have also used a single-period newsvendor problem to study the transaction that finances inventory. In contrast to our model, they assume that the cost of capital of the bank and the newsvendor are both fixed at the risk-free rate, which is most appropriate when markets are efficient and complete. Consequently, as discussed immediately after Proposition 13.2 in Section 13.2.2, this optimal order quantity does not depend on the CCNV's level of equity. Their single-period model is further enriched by including the impact of taxes and bankruptcy costs. Kouvelis and Zhao (2008) also assume a complete and perfect financial market and therefore reach the same conclusion as Xu and Birge (2004) that the operational and financial decisions of a firm can be decoupled.

Xu and Birge (2004) assume perfectly competitive loan market such the creditor chooses the interest rate to generate a fixed expected rate of return, the risk-free interest rate. In contrast, a key feature of our model, as in Buzacott and Zhang (2004), is that the bank is a monopolist, setting the interest rate to maximize its expected profit. Moreover, as in the work of Li et al. (1997), the newsvendor's problem may be modeled as a multiperiod problem that explicitly examines the cost of reorganization when bankruptcy risks are significant. Our single-period model could be used as a building block for considering such models when liquidity or working capital is an issue.

Importantly, our work and Xu and Birge (2004) may be viewed as single period instances of the dynamic aspects of asset management financing originally considered by Buzacott and Zhang (2004). In particular, our Proposition 13.1 in Section 13.2 is a distilled version of their Theorem 3. While some of our analysis has a similar flavor to theirs, the key distinction is that we formally treat the bank as the Stackelberg leader and prove that the Stackelberg equilibrium is unique. In contrast, there are many retailers in Buzacott and Zhang (2004). Their bank behaves more like a monopolist who offers take-or-leave-it loans to retailers with different internal capitals.

Xu and Zhang (2006) allow a supplier to offer loans and provide goods to its capital constrained retailers in a monopoly setting. They examine whether the typical supply chain contracts could achieve channel coordination. Zhou and Groenevelt (2007) also discuss supply chain coordination in the presence of supplier financing. They also briefly discuss financing by profit-seeking banks. In contrast, our bank provides the newsvendor a loan but not physical goods. We show that a nonlinear loan schedule can coordinate the channel.

We model the problem faced by the CCNV in Section 13.2.1 and the bank's problem in Section 13.2.2. Some comparative statics of the equilibrium are presented in Section 13.2.3. In Section 13.3, we provide numerical studies. In Section 13.4, we propose a nonlinear loan schedule to achieve channel coordination. Finally, Section 13.5 concludes with remarks.

13.2 The Stackelberg Game

13.2.1 THE NEWSVENDOR'S PROBLEM

The newsvendor places an order for Q units at unit cost c before the selling season starts and sells them at unit price p, where $(p > c)$. The seasonal demand D follows a cumulative probability distribution $F(\cdot)$ and density probability distribution $f(\cdot)$ on R^+. Define $\overline{F}(x) = 1 - F(x)$. Conveniently assuming a zero salvage value for any excess inventory and zero goodwill costs for lost sales, the classical newsvendor's optimal order quantity, denoted by Q_0, satisfies $\overline{F}(Q_0) = c/p$. However, since the newsvendor is *capital-constrained* (i.e., its internal capital; henceforth addressed as *equity* to be differentiated from debt), represented by η, is less than cQ_0, the CCNV may find it profitable to finance additional procurement by borrowing B from a financial intermediary, such as a bank, at an interest rate, r. The CCNV borrows just enough to purchase the desired quantity Q at the financing rate r, thereby $B = cQ - \eta$. To secure the loan, the CCNV pledges the initial $(1 + r)B$ of its revenue to the bank and receives the residual revenue, if any, after sales revenue realizes. Thus, the CCNV's optimization problem is:

$$\max_{Q \geq \eta/c} \pi_n = -\eta - B(1+r)\overline{F}\left(\frac{B(1+r)}{p}\right) + p\int_{\frac{B(1+r)}{p}}^{Q} x dF(x) + pQ\overline{F}(Q)$$

It is insightful and practical to make the following changes of variables:

$$m = \frac{p}{1+r} \qquad y = \frac{(cQ - \eta)(1+r)}{p} = \frac{B}{m} \tag{13.1}$$

Note that m represents the discounted revenue from one unit of sales and that y is the number of units that must be sold to fully payoff the loan B taken by the CCNV. Hence, the CCNV defaults with probability $F(y)$.

Using (13.1), the CCNV's objective function becomes:

$$\begin{aligned}\pi_n &= -\eta + p\int_y^Q (x-y)dF(x) + p(Q-y)\int_Q^\infty dF(x) \\ &= -\eta + p\int_y^Q \overline{F}(x)dx\end{aligned} \tag{13.2}$$

In contrast to the formulation for the unconstrained newsvendor, the lower limit, y, of the integral captures the revenue that is in excess of the amount pledged to the bank and the first term recognizes the procurement that is financed by equity. Our formulation of the newsvendor's problem is a variation of the retailer's problem considered by Buzacott and Zhang (2004; Theorem 3, p. 1283). They have shown that if the demand distribution has increasing failure rate (IFR), the optimal solution can be fully characterized by manipulating the Karush-Kuhn-Tucker conditions to yield:

Proposition 13.1 *For any $\eta < cQ_0$, if the demand distribution function, $F(\cdot)$, is IFR (has an increasing failure rate), then the newsvendor's optimal ordering quantity, Q^*, is uniquely determined as follows:*

$$Q^* = \begin{cases} \frac{\eta}{c} & \text{if } \eta/c > \overline{F}^{-1}\left(\frac{c}{m}\right) \\ \hat{Q}, & \text{otherwise,} \end{cases} \tag{13.3}$$

where

$$m\overline{F}(\hat{Q}) = c\overline{F}(y) \tag{13.4}$$

Proof of all propositions are attached in the Appendix at the end of the chapter. In the first case of (13.3), the CCNV exhausts its own equity η but does not borrow from the bank. This occurs when the interest rate is too high, suggesting that for every η there is an upper limit on the rate that the bank may charge. It also readily follows that this upper limit is decreasing in the CCNV's equity. Alternatively, if the interest rate is sufficiently low, after exhausting its equity, the CCNV would seek additional financing to order $\hat{Q}$, which uniquely satisfies (13.4).

Given the CCNV's best response to the interest rate set by the bank, we will next consider the bank's problem. As the Stackelberg leader, it must take into account how the CCNV will respond to a change in the interest rate r or m. This response, from applying the Implicit Function Theorem to (13.4), is given by:

$$\frac{dy^*}{dm} = \frac{c\overline{F}[(\eta+my)/c] - myf[(\eta+my)/c]}{m^2 f[(\eta+my)/c] - c^2 f(y)}. \tag{13.5}$$

The sign of the right-hand side of above equation cannot be determined. Numerical examples show that it can be both ways. However, we show in the next section that at the equilibrium point y^* is monotone in m.

13.2.2 THE BANK'S PROBLEM

At the second stage of the Stackelberg game, the CCNV decides how much to borrow given the interest rate r charged by the bank. Thus, at the first stage, anticipating the newsvendor's response, the bank determines r or m to maximize its expected profit. This yields as counterpart to (13.2):

$$\max_{r} \pi_b = B(1+r)\overline{F}(y) + p\int_0^y x dF(x) - B \tag{13.6}$$

The first term of (13.6) is the interest plus the principle that the bank receives when the CCNV does not default, while the second term is the expected sales revenue when the newsvendor defaults. It follows from Proposition 13.1 that the CCNV does not borrow if $m < c/\overline{F}(\eta/c)$. In order to induce the CCNV to borrow, the bank chooses m between $c/\overline{F}(\eta/c) \leq m < p$ where $m < p$ is equivalent to $r > 0$. Using (13.1), (13.6) can be rewritten as:

$$\max_{c/\overline{F}(\eta/c) \leq m < p} \pi_b = p\int_0^y \overline{F}(x)dx - my \tag{13.7}$$

To understand the bank's problem, one way to proceed is to take the partial derivative of π_b with respect to y and compute the bank's desired y. Analogous to (13.11) in Buzacott and Zhang (2004), the first-order condition of π_b with respect to y yields $\overline{F}(y) = m/p$. However, such an approach fails to recognize that the CCNV responds to m by choosing Q or y. This strategic interaction, is captured by the total derivative of π_b with respect to m, which yields the first-order condition:

$$\begin{aligned} \frac{d\pi_b}{dm} &= -y^*(m) + \frac{\partial \pi_b}{\partial y}\frac{dy^*}{dm} \\ &= -y^*(m) + [-m + p\overline{F}(y^*(m))]\frac{dy^*}{dm} = 0 \end{aligned} \tag{13.8}$$

Note that dy^*/dm is given by (13.5). A rational bank would select m in a range so thats (13.4) of Proposition 13.1 that in the admissible range of m, $\frac{\partial \pi_b}{\partial y} \geq 0$. That is, the bank would select the loan size such that its expected profit is nondecreasing in the size of the loan. This allows us to conclude that the bank will select m such that $m - p\overline{F}(y^*) < 0$. As a consequence, we can conclude immediately that at equilibrium $dy^*/dm > 0$. This monotonicity guarantees that for any $y*$, there is only one corresponding m. In addition, by Proposition 13.1, for every m, the retailer's best response y is unique. Therefore, the equilibrium point of the Stackelberg game is unique.

Moreover, using (13.4),

$$\bar{F}(\hat{Q}) = c\bar{F}(y)/m > \frac{c}{m} \times \frac{m}{p} = c/p$$

This implies that $\bar{F}(\hat{Q}) > \bar{F}(Q_0)$. Therefore, in equilibrium the CCNV would produce less than the first-best quantity. These conclusions are summarized as follows.

Proposition 13.2 *If $F(\cdot)$ is IFR, (a) the Stackelberg game between the bank and the newsvendor has a unique equilibrium point (y^*, m^*), which satisfies (13.3), (13.5), and (13.8), and (b) $0 < y^* < \overline{F}^{-1}(m^*/p)$, $\overline{F}^{-1}(c/m^*) < \hat{Q} < Q_0$ and $0 < dy^*/dm$.*

Part (b) of Proposition 13.2 is in contrast to Xu and Birge (2004) who assume that both the bank and the CCNV expect a return at the risk-free rate r_f, resulting in $\hat{Q} = Q_0 = \overline{F}^{-1}(c/p(1 + r_f))$. Here the term $(1 + r_f)$ reflects an adjustment to p for the time value for money, so it suggests a decoupling of procurement and financing decisions. In contrast, we have just demonstrated that the procurement decision is directly linked with financial decisions. Moreover, at equilibrium the CCNV purchases less than would a traditional newsvendor when the capital market is not competitive (i.e., the bank is strategic even in the absence of taxes and bankruptcy costs).

13.2.3 SOME COMPARATIVE STATICS

Having established that the equilibrium is unique, we are in a position to perform comparative statics. This yields the following comparative statics with respect to η, p, and c.

Proposition 13.3 *(a) As η increases, y^*, B^*, and r^* decrease, m^* increases. (b) For a given η, as p increases, y^*, B^*, $\hat{Q}$, r^*, and m^* increase. And (c) for a given η, as c increases, y^* increases, but $\hat{Q}$, m^* and B^* decrease.*

If the CCNV has more equity, it needs to borrow less. Hence, to induce the newsvendor to borrow, the bank has to lower its interest rate, equivalently, raise m. Therefore, B^* and y^* are lower if η is greater. However, surprisingly, the order quantity $\hat{Q}$ is not necessarily increasing in η. To examine this, applying the Implicit Function Theorem on (13.4) yields:

$$\frac{d\hat{Q}}{d\eta} = \frac{\overline{F}(\hat{Q})\frac{dm^*}{d\eta} + f(y^*)\frac{dy^*}{d\eta}}{m^* f(\hat{Q})}$$

Because $dm^*/d\eta > 0$ and $dy^*/d\eta < 0$, the sign of $d\hat{Q}/d\eta$ is not invariant, as can be confirmed by numerical examples.

For a fixed η, a higher p implies that the compensation that the bank receives is greater when the newsvendor defaults, while its revenue is fixed at $B(1+r)$ otherwise. Hence, the bank is willing to offer a lower r (higher m). Since y^* is increasing in m at equilibrium, y^* must be also increasing in p. So is B^* since $B^* = m^*y^*$. The interest rate that the bank charges, however, is increasing in p because the absolute value of the marginal decreasing rate of m with respect to p is less than one and $r = p/m - 1$. Not surprisingly, both the newsvendor and the bank benefit from a higher unit selling price of the product since the expected profits for both are increasing in p.

For a fixed η, a higher c yields a lower profit margin, thereby a higher threshold above which the CCNV survives, (i.e., a greater y^*). Consequently, the bank charges a higher r (lower m), and the CCNV borrows and purchases less.

Although the uniqueness of equilibrium is easily established, explicit expressions for (y^*, m^*) are not easily available. In order to obtain additional insights, we conducted a variety of numerical studies, which are reported in next section.

13.3 A Numerical Study

To further explore the sensitivity of the decisions made by the CCNV and the bank, we conduct a computational study, showing that generally, as we would expect, as the demand becomes less certain (as captured by the coefficient of variation), overall performance improves. Demand distributions are drawn from one of three members of the Weibull family with $\beta = 1, 2, 3$ and $\alpha = 1, (\Gamma(3/2))^{-1}, (\Gamma(4/3))^{-1}$ so that the mean is standardized at 1.[1] Consequently, as β increases from 1 (the exponential case) to 2 to 3, skewness decreases, making the distribution more and more symmetric.

Moreover, we set $p = 2$, $c = 1$. Thus, $F(Q_0) = (p - c)/p = 0.5$, where $Q_0 = 0.6913, 0.9394, 0.9911$, respectively, for $\beta = 1, 2, 3$. The respective optimal expected profits are 0.30685, 0.5825, and 0.7059.

Since the CCNV is limited by its capital, let $0 < \eta < cQ_0$. In the computational results we report how the operating decisions, performance measures, and efficiency vary with η. Some technical details on the computations are in the Appendix at the end of the chapter.

We begin by examining the decisions of the CCNV, which are reported in Figure 13.1. The first panel shows that as η increases, the CCNV finances less of its order from the bank. However, less intuitively, this decrease is not regular enough to result in a monotone decrease in the total order quantity. Moreover, for a fixed η, the principal B or the order quantity Q^* is ranked by β, which in a sense captures decreasing risk reflected in decreasing skewness.

We now turn to the decision made by the bank who sets m or equivalently r. The first panel of Figure 13.2 indicates that as η rises, r (m) decreases (increases). A greater η means that the CCNV is less capital-constrained. Thus, the bank is

[1] The Weibull distribution function has the form $F(x) = 1 - e^{-(\frac{x}{\alpha})^\beta}$ for $x > 0$.

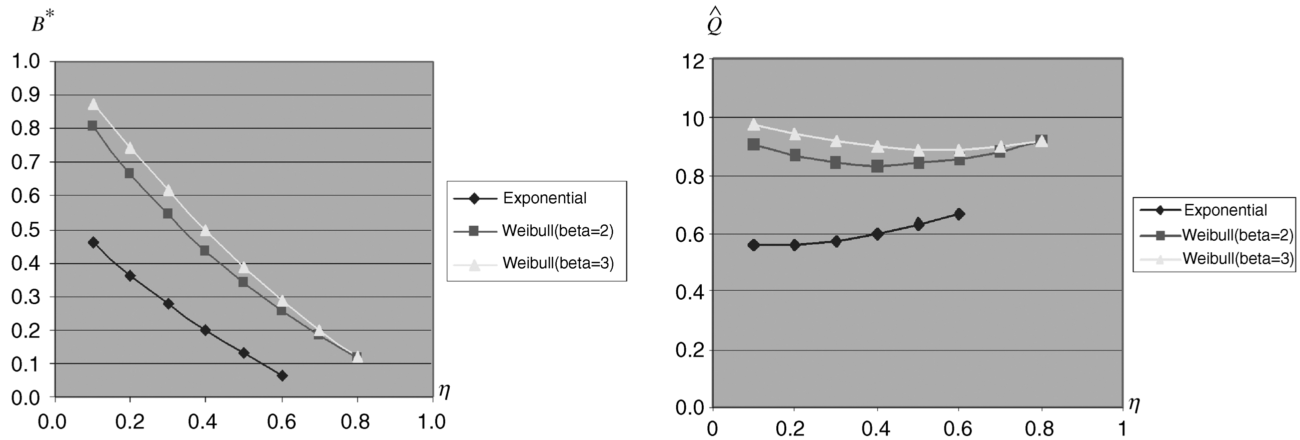


FIGURE 13.1 The effect of equity on CCNV.

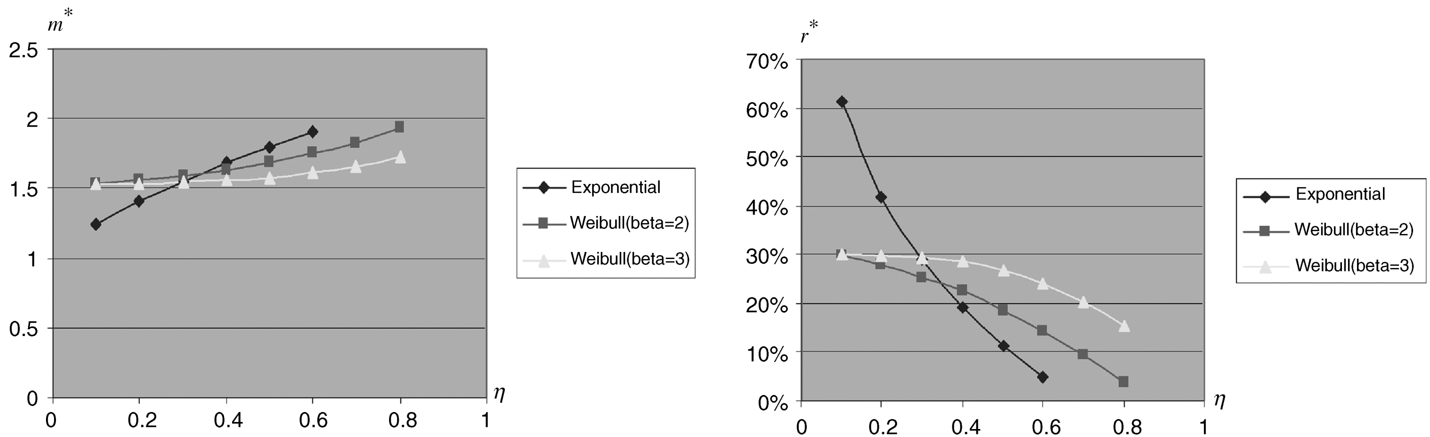


FIGURE 13.2 The effect of equity on the bank.

compelled to lower r (raise m) to induce more borrowing. However, in contrast to Figure 13.1, there is no clear relationship in the ranking by β for a fixed η.

As η increases the optimal payment drops, which is captured by y^* in the first panel of Figure 13.3. Furthermore, for a fixed η, y^* is ranked by β. As skewness decreases, increasingly greater amounts are borrowed. However, the survival probability presented in the second panel of Figure 13.3 does not capture this regularity possibly because the Weibull distributions as used here do not exhibit first-order stochastic dominance in β.

Having examined the operating decisions made by the two parties, we now turn to the economic impacts. We observe from Figure 13.4 that for a given demand distribution, the expected profits are monotone in η. The CCNV's profit is increasing in η, but the converse is true for the bank's profit. Moreover, the CCNV's profit is ranked by β, and the bank's is almost consistently ranked by β, except at relative tight capital constraints.

Furthermore, as seen in the left panel of Figure 13.5, the CCNV has a falling return on investment as η increases. This is not surprising since economic intuition suggests that capital should have diminishing returns. Interestingly, however, this economic intuition is not quite borne out at the bank whose ROI (return on investment) seems increasing and then decreasing in η. This counterintuitive result may be mitigated by noting that ROI on the total investment can be verified to be decreasing in η. In addition, from Figure 13.6, the channel's ROI, defined as $ROI_w = (\pi_n^* + \pi_b^*)/(cQ^*)$, is ranked by β, and it appears that there is an η at which ROI_w reaches its maximum.

By Proposition 13.2, the CCNV procures a smaller order quantity than if it were not capital-constrained. Thus, there is a loss of economic efficiency. To capture this loss of efficiency, we first examine the value of borrowing to the CCNV. Define $E_b = (\pi_n^* - \pi_{nb})/\pi_{nb}$ where π_{nb} is the CCNV's expected profit without borrowing. Consistent with the findings on ROI, the CCNV enjoys relatively less benefit from borrowing as it becomes less capital-constrained; it also benefits relatively more when the distribution is less skewed. Hence, the ordering of incremental benefit of borrowing by β is quite intuitive as seen clearly in the first panel of Figure 13.7.

Finally, we examine the loss of economic efficiency in the financial channel's profit relative to the unconstrained case. Define channel efficiency as $E_c = (\pi_n^* + \pi_b^*)/\pi^*$ where π^* is the optimal expected profit of an unconstrained newsvendor. The channel efficiency exhibits multimodal patterns from the second panel of Figure 13.7. The channel efficiency tends to increase in η but has substantial fluctuations. However, it is more stable for $\beta = 3$ than for the other cases, suggesting that as distributions become more symmetric, the channel efficiency becomes relatively invariant. However, we must qualify this observation since we did not find this conjecture to hold if demand is uniformly distributed. The results for this case are presented in the Appendix at the end of the chapter. In fact, we find that the case of uniform demand qualitatively behaves more like the exponential case than like the other Weibull distributions.

The CCNV does not choose the first-best order quantity because it reacts to a given interest rate. If instead the bank sets a loan schedule that allows the

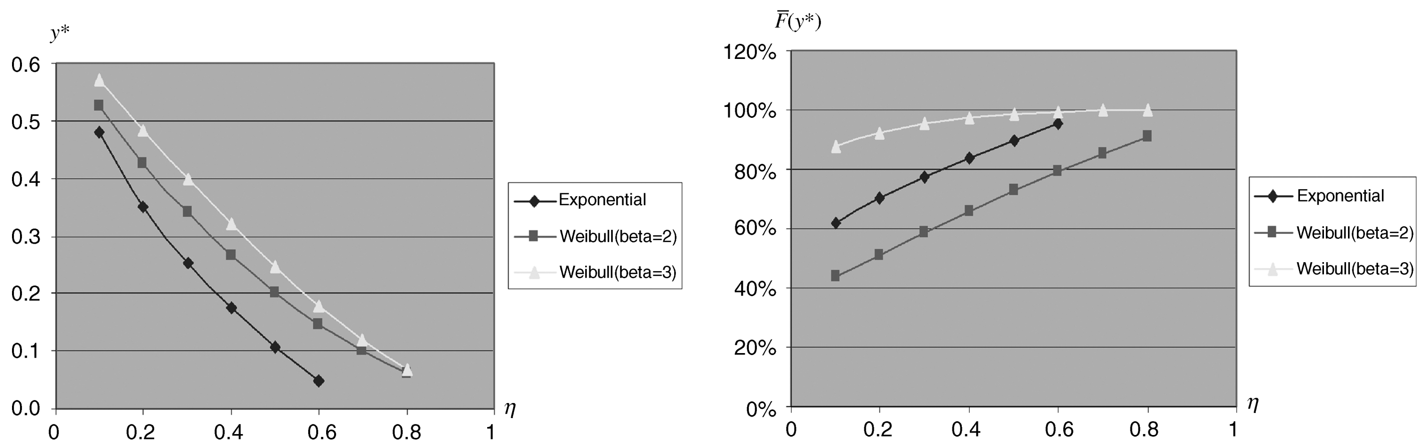


FIGURE 13.3 The effect of equity on default probability.

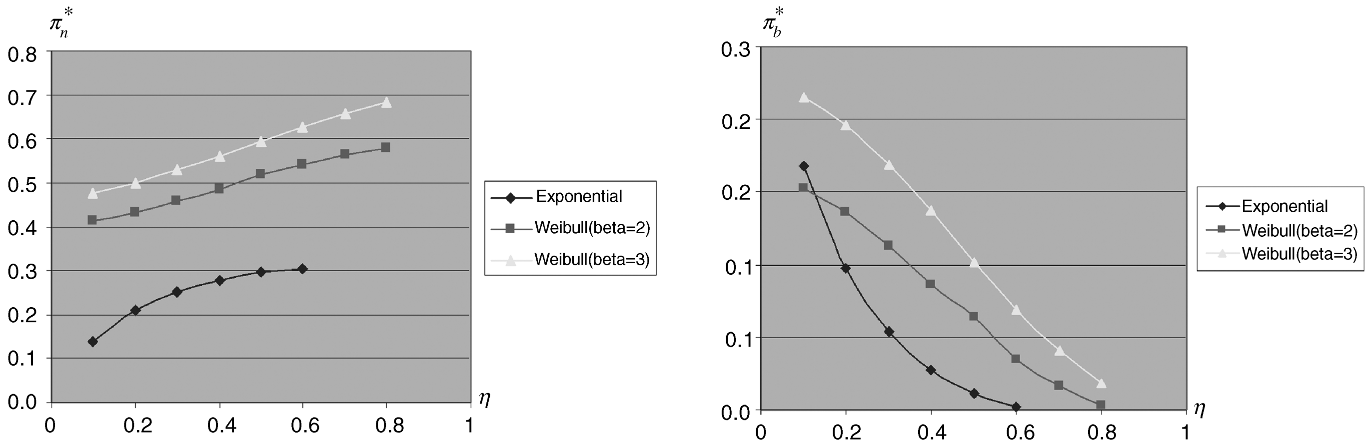


FIGURE 13.4 The effect of equity on profits.

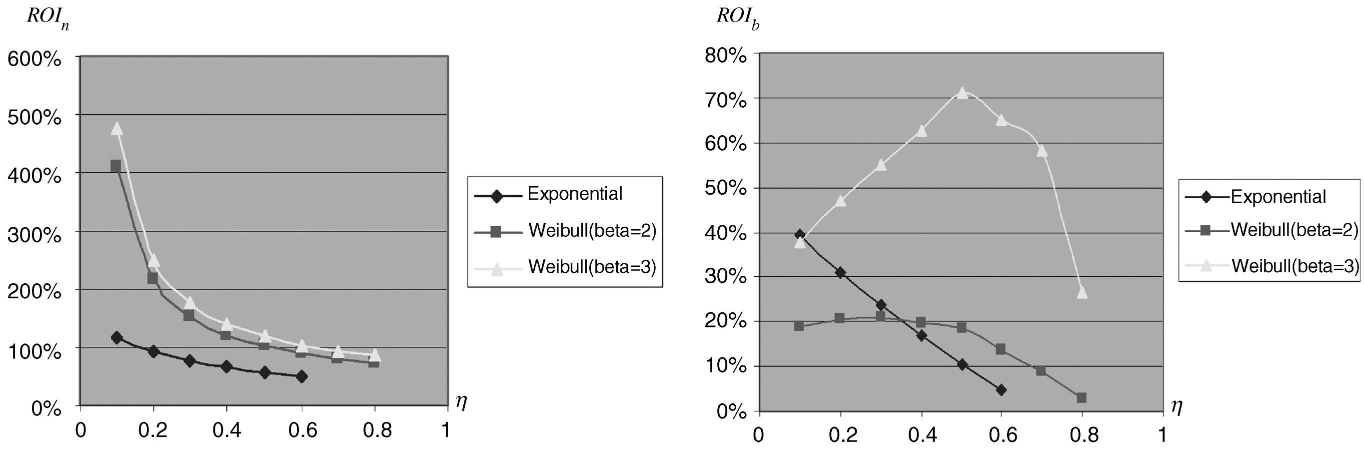


FIGURE 13.5 The effect of equity on ROI.

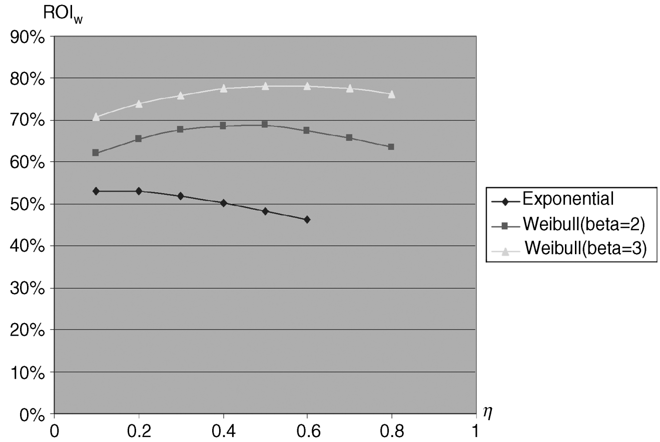


FIGURE 13.6 The effect of equity on channel's ROI.

CCNV to choose both the debt and the interest rate from an appropriate class of nonlinear loan schedules, then the equilibrium yields a first-best solution as is shown next.

13.4 Coordinating Loan Schedules

Thus far we have assumed that if the CCNV does not obtain additional financing, it can only order $Q_\eta = \eta/c$, units yielding an expected profit of π_n^η. If the CCNV borrows at equilibrium, it borrows sufficient funds to procure an amount somewhat more than η/c but less than the channel optimal quantity Q_0, which yields the optimal expected channel profit π_0. Thus, at equilibrium there is some loss of channel efficiency. This loss of efficiency arises because we have assumed that the CCNV responds to the rate r or m set by the bank by choosing a loan amount B or y. In this section, we assume that the bank proposes an appropriate nonlinear loan schedule $r(B)$ or equivalently $m(y)$ to induce the CCNV to order the channel optimal quantity Q_0. Since now m is a function of y, if the CCNV borrows, then the first-order condition of π_n with respect to y yields:

$$\overline{F}(Q)\left(m + y\frac{dm}{dy}\right) = c\overline{F}(y)$$

For this schedule to achieve coordination $\overline{F}(Q) = c/p$ must hold. Consequently:

$$m + y\frac{dm}{dy} = \frac{dB}{dy} = p\overline{F}(y) \tag{13.9}$$

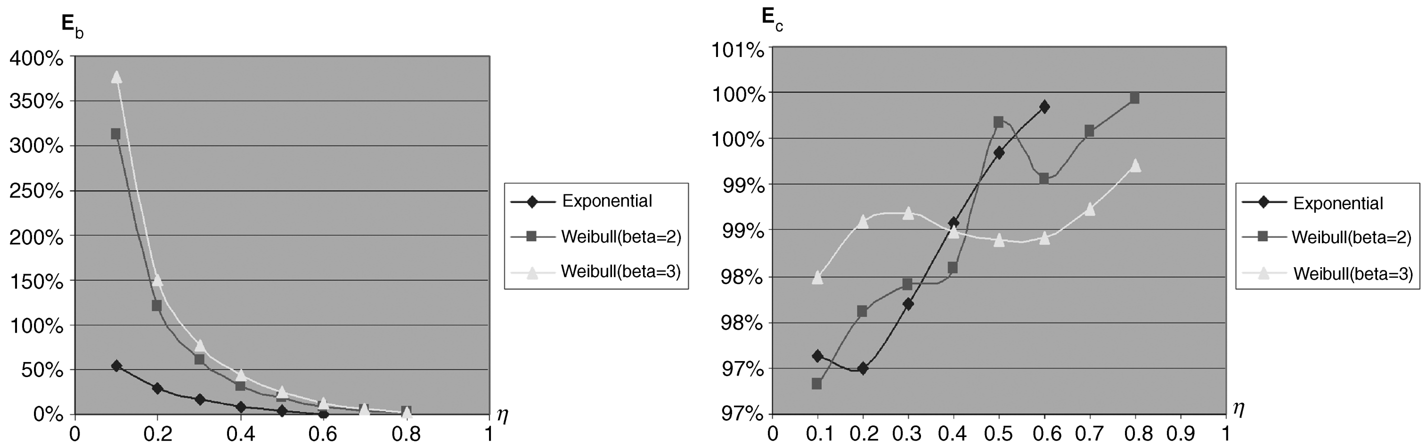


FIGURE 13.7 The effect of equity on efficiency.

Define:

$$g(y) = \int_0^y \overline{F}(x)dx$$

Then (13.9) becomes:

$$B = cQ_0 - \eta = my = pg(y) - K \tag{13.10}$$

The range and the meaning of K will be specified later. Hence, at equilibrium, π_n^*, the CCNV's optimal expected profit is:

$$\begin{aligned}\pi_n^* &= -\eta + p\int_y^{Q_0} \overline{F}(x)dx \\ &= -\eta + p[g(Q_0) - g(y)] \\ &= -\eta + pg(Q_0) - my - K = \pi_0 - K,\end{aligned} \tag{13.11}$$

By construction, $\pi_0 = pg(Q_0) - \eta - my$, and, π_b^*, the bank's expected profit is:

$$\pi_b^* = p\int_0^y \overline{F}(x)dx - my = pg(y) - my = K \tag{13.12}$$

Note that my is the loan (i.e., the bank's cost), and $pg(y)$ is the expected cash flow that the bank receives when the CCNV defaults. Adding (13.11) and (13.12) yields the expected total channel profit:

$$\pi_n^* + \pi_b^* = p\int_0^Q \overline{F}(x)dx - cQ_0 = pg(Q_0) - cQ_0 = \pi_0 \tag{13.13}$$

Note that $cQ_0 = \eta + my$. Define π_n^η as the CCNV's expected profit without borrowing:

$$\pi_n^\eta = -\eta + p\int_0^{\eta/c} \overline{F}(x)dx = -\eta + pg(\eta/c) < \pi_0. \tag{13.14}$$

It is clear now that K is the share of expected profit that the bank enjoys from the transaction. Furthermore, since π_n^η is increasing in η for $0 < \eta < cQ_0$ and π_0 is fixed, for any transaction between the two parties to occur and for the bank to have incentives to achieve coordination, $0 \le K \le \pi_0 - \pi_n^\eta$. That is, the CCNV must not be worse off by borrowing the loan. Since π_n^η is increasing in η, the upper bound on K is decreasing in η. If $K = \pi_0 - \pi_n^\eta$, then the bank extracts all the surplus generated by the loan. If $K = 0$, then $\pi_b^* = pg(y) - my = 0$, and the CCNV retains all the benefits originated from borrowing. How the surplus generated by the loan is shared between the players depends on the relative power of the bank and the CCNV and the bargaining process. In addition, analogous to the well-known channel coordinating two-part tariff, we could interpret m as the price at which the bank offers the CCNV the loan $cQ_0 - \eta$ and it breakseven, and K as the expected fee that it charges the CCNV to extract a profit from lending. The results are summarized as follows.

TABLE 13.1 Coordination Value through Nonlinear Loan Schedule

η	π_b	π_n^η	$\pi_n^\eta + \pi_b$	$\frac{\pi_0-(\pi_b+\pi_n^\eta)}{\pi_0}$
0.1	0.3232	0.1447	0.4679	6.41%
0.2	0.2349	0.2309	0.4658	6.84%
0.3	0.1698	0.2995	0.4694	6.13%
0.4	0.1194	0.3556	0.4750	4.99%
0.5	0.0800	0.4013	0.4813	3.74%
0.6	0.0497	0.4376	0.4873	2.53%
0.7	0.0273	0.4653	0.4925	1.49%
0.8	0.0119	0.4847	0.4965	0.69%
0.9	0.0029	0.4962	0.4992	0.18%

Proposition 13.4 *If the bank offers a loan schedule* $cQ_0 - \eta = p\int_0^y \overline{F}(x)\, dx - K$*, then* $\pi_b^* = K$*,* $\pi_n^* = \pi_0 - K$ *for* $0 < K < \pi_0 - \pi_n^\eta$ *and* $0 < \eta < cQ_0$*, and the channel achieves coordination.*

In addition, under the nonlinear loan schedule proposed above, it is easy to show that the comparative statics of Proposition 13.3 continue to hold. However, by design $Q^* = Q_0$, the channel's first-best quantity, so it is invariant to η. And, the channel efficiency is 100%, so the multimodal patterns found in our numerical examples cannot arise.

We now illustrate above coordinating schedule with two examples. Let $p = 2$ and $c = 1$. First, let demand be uniformly distributed between 0 and 2. So $Q_0 = 1$, $\pi_0 = 1$, and $g(y) = y/2 - y^2/4$. Hence, $0.5 - \eta = y - y^2/2 - K$. Solving the quadratic equation yields the loan schedule $y^* = 1 + \sqrt{2\eta - 2K}$ where $0 < K < 0.5 - \eta + \eta^2/2$ and $0 < \eta < 1$. Table 13.1 shows the value of coordination for the channel using the nonlinear loan schedule provided above. The value of coordination is not necessarily monotone in η. Overall, it provides the least value when the CCNV is not so capital-constrained and it has funds to finance near the first-best level (e.g., when $\eta = 0.9$ in Table 13.1). When the CCNV is severely constrained, the value of coordination is as high as nearly 7% of the first-best channel profit.

Second, if demand follows the exponential distribution with mean 1, then $Q_0 = ln2$, $\pi_0 = 1 - ln2$, $g(y) = 1 - e^{-y}$, and $\pi_n^\eta = 2 - \eta - 2e^{-\eta/c}$. For $0 < \eta < ln2$, the loan schedule is $y^* = -ln(2 - ln2 - K + \eta)$ where $0 < K < \pi_0 - \pi_n^\eta = \eta + 2e^{-\eta} - 1 - ln2$ and $0 < \eta < ln2$.

13.5 Concluding Remarks

We examine how optimal inventory decisions are affected when an organization has limited availability of working capital. To glean sharp insights, we used the fundamental case of the newsvendor model. We first focus on a Stackelberg game in which the newsvendor is the follower and the bank is the leader. This setting captures instances when specialized knowledge precludes existence of market efficiency. Our results show that if the cost of borrowing is not too high, the capital-constrained newsvendor borrows funds to procure an amount that is somewhat less than would be ideal. In return the bank charges an interest rate that decreases in the equity position of the CCNV. We also report that holding the mean fixed, as demand becomes less skewed, the financial channel's profits tend to rise, and the equilibrium leads to relatively less loss in channel efficiency.

A key feature of our model is that the interest rate to be charged is determined by the bank. In contrast, Xu and Birge (2004) consider a model in which the newsvendor, while capital constrained, like the bank, has a cost of capital that is the risk-free rate. In this case, since the capital market is competitive and thus the bank is nonstrategic, they find that the operating decision is independent of the financing decisions made by their newsvendor. Their single-period model is further enriched by including the impact of taxes and bankruptcy costs. Moreover, as in the work of Li et al. (1997), the newsvendor's problem may be modeled as a multiperiod problem that explicitly examines the cost of reorganization when bankruptcy risks are significant. Our single-period model could be used as a building block for considering such models when liquidity or working capital is an issue.

Like most games, our model yields an equilibrium that has some loss of efficiency. Our numerical work suggests that efficiency is relatively high and tends to increase with decreasing skewness of the demand distribution. However, even this loss in efficiency can be eliminated if a profit-splitting nonlinear loan mechanism is invoked because it always yields the first-best solution. Hence, as in Xu and Birge (2004), bankruptcy costs and taxes can be incorporated seamlessly. Moreover, the model can be also seamlessly adapted if the bank requires fixed assets such as fixed capital and real estate from the CCNV as collateral of the loan.

More importantly, under our alternative specification, the first-best solution is reached, which is the equilibrium in the financial channel between the newsvendor and the bank. Consequently, it is interesting to ask, as have Xu and Zhang (2006), whether a first best solution can be found in which the newsvendor is capital-constrained and the channel consists of it and its supplier who has a constant unit production cost. Our results raise the issue of whether the first-best solution exists for this case as well.

Indeed, the first-best solution exists for this more complex case if the problem is decoupled into two strategic games. In one game, exactly as in the alternative specification, the newsvendor obtains financing from an intermediary or its supplier. And, in the other game, a variety of channel coordination mechanisms, like those described by Cachon (2005) may be used to obtain the first-best solution.

Finally, if the newsvendor and bank embark on a Nash game, that is, the bank loses its leadership position, then the game degenerates because the bank would not lend and the CCNV would not borrow. To see this, from (13.7),

$$\frac{d\pi_b}{m} = -y \leq 0$$

Compared with (13.8), above equation misses the term that captures the bank's first-mover advantage. As a result, the bank would set m as small as possible (i.e., $m = 0$). By Proposition 13.1, the CCNV would not borrow if $m < \frac{c}{\overline{F}(\eta/c)}$. Hence, under bankruptcy risk, the bank may not be able to recoup the loan principle, the Stackelberg setting, which awards the bank a first-over advantage, seems necessary and fits well with reality since banks often have the power in determining whether or not to extend credits to small businesses.

APPENDIX

Proof of Proposition 13.1

Due to the unimodality of π_n in Q, the CCNV's problem can be solved by invoking the Kraush-Kukh-Tucker conditions:

$$\begin{aligned} &m\overline{F}(Q) - \overline{F}(y) + \lambda = 0 \qquad \lambda \geq 0 \\ &\lambda(\eta/c - Q) = 0 \qquad \eta/c - Q \leq 0 \end{aligned}$$

Solving the above equations yields Proposition 13.1.

Proof of Proposition 13.2

For part (a), due to the unimodularity of π_n in m, there always exists a unique y^* for each m.

For part (b), $\overline{F}(y) < 1$ and (13.4) yields $Q^* > \overline{F}^{-1}(\frac{c}{m})$. If $Q^* > \overline{F}^{-1}(\frac{c}{p})$, then $\overline{F}(Q^*) > c/p$. Therefore:

$$\overline{F}(Q) = c\overline{F}(y)/m \leq c/p$$

Thus, $\overline{F}(y) < m/p$, so $\partial\pi_b/\partial y < 0$, which cannot happen for a profit-maximizing bank.

Proof of Proposition 13.3

Applying the Implicit Function Theorem on (13.4) yields:

$$\frac{dm}{d\eta} = \frac{mf(Q)}{c\overline{F}(Q) - myf(Q)}$$

Since:

$$\frac{dy}{dm} = \frac{c\overline{F}(Q) - myf(Q)}{f(y) - m^2 f(Q)} > 0$$

At equilibrium, $c\overline{F}(Q) - myf(Q) > 0$. Therefore, $dm/d\eta > 0$. Consequently, $dy/d\eta < 0$ and $dB/d\eta < 0$. Similarly, applying the Implicit Function Theorem on (13.8) yields:

$$\frac{dy}{dp} = \overline{F}(y)\frac{dy}{dm} > 0 \qquad \frac{dm}{dp} = \overline{F}(y) > 0$$

For part (c), because $dy/dm > 0$ at (y^*, m^*), $B^* = m^*y^*$, and $Q^* = \eta + B^*$, the monotone property of Q^* with respect to p is immediate. Substituting $p/(1+r)$ for m in (13.8) and applying the Implicit Function Theorem yields:

$$\left[-\frac{1}{1+r} + \overline{F}(y)\right]\frac{dy}{dm}dp - \frac{p}{(1+r)^2}\frac{dy}{dm}dr = 0$$

Hence, since $\overline{F}(y) > 1/(1+r)$:

$$\frac{dr}{dp} = \frac{\overline{F}(y) - \frac{1}{1+r}}{\frac{p}{(1+r)^2}} > 0$$

Another way to prove $dr/dp > 0$ is to observe that since $dm/dp = \overline{F}(y) < 1$ and $r = p/m - 1$ because $p\overline{F}(y) > m$:

$$\frac{dr}{dp} = \frac{m - p\overline{F}(y)}{m^2} > 0$$

The monotone properties of y^*, Q^*, B^*, and m^* with respect to c can be proved in the same fashion.

Computation When Demand Follows Weibull Distributions

Now (13.4) and (13.8) can be specialized as:

$$y^\beta = (\eta + my)^\beta - \alpha^\beta \ln(m) \tag{13.15}$$

$$-y + [m + 2e^{-(\frac{y}{\alpha})^\beta}]\frac{dy}{dm} = 0 \tag{13.16}$$

A necessary and sufficient condition for $y > 0$ is $\ln m > (\frac{\eta}{\alpha})^\beta$. The unique solutions for (13.15) and (13.16) can be found by a straightforward line search.

Computation When Demand Follows a Uniform Distribution

Let demand D follow a uniform distribution in $[0, 2]$. If the newsvendor is not capital-constrained, then $Q^* = Q_0 = 1$ and $\pi_n^* = 0.5$. Equations (13.4) and (13.5) can be specialized as:

$$y = \frac{2m - \eta m - 2}{(m^2 - 1)}$$

$$\frac{dy}{dm} = \frac{(\eta - 2)m^2 + 4m + (\eta - 2)}{(m^2 - 1)^2} \tag{13.17}$$

A necessary and sufficient condition for $y > 0$, so that the newsvendor borrows, is $m > 2/(2 - \eta)$. Using (13.17), (13.8) becomes

$$\frac{2m - \eta m - 2}{m^2 - 1} + \left(m - 2 + \frac{2m - 2 - \eta m}{m^2 - 1}\right) \frac{(\eta - 2)(m^2 + 1) + 4m}{(m^2 - 1)^2} = 0 \tag{13.18}$$

For each η, solving the above polynomial equation yields $m^* > 1/(2 - \eta)$. Subsequently, y^* and the respective expected profits of the newsvendor and the bank can be calculated.

REFERENCES

Archibald T., Thomas L., Betts J., Johnston R. 2002. Should start-up companies be cautious? inventory policies which maximize surviral probabilities, *Management Science,* **48**:1161–1174.

Babich V., Sobel M.J. 2004. Pre-ipo operational and financial decisions. *Management Science,* **50**:935–948.

Birge J.R. 2000. Option methods for incorporating risk into linear capacity planning models. *Manufacturing and Service Operations Management,* **2**:19–31.

Bounds G. 2006. Pen maker struggles with too much demand. *The Wall Street Journal,* 14.

Boyabatlih O., Toktay L.B. 2010. Stochastic capacity investment and flexible versus dedicated technology choice in imperfect capital markets. Working paper. Singapore: Singapore Management University.

Buzacott J.A., Zhang R. 2004. Inventory management with asset-based financing. *Management Science*, **24**:1274–1292.

Cachon G. 2005. Supply chain coordination with contracts In: Graves S.C., de Kok T., editors, Handbooks in Operations Research and Management Science. New York: Elsevier Science, North-Holland.

Gaur V., Seshadri S. 2005. Hedging inventory risk through market instruments. *Manufacturing and Service Operations Management,* **7**:103–120.

Gupta D., Wang L. 2009. A stochastic inventory model with trade credit. *Manufacturing and Service Operations Management,* **11**:4–18.

Huchzermeier A., Cohen M.A. 1996. Valuing operational flexibility under exchange rate uncertainty. *Operations Research,* **44**:100–113.

Kogut B., Kulatilaka N. 1994. Operational flexibility, global manufacturing, and the option value of a multinational network. *Management Science* **40**:123–139.

Kouvelis P., Li C.-L. 1999. Flexible and risk sharing supply chain contract under price uncertainty. *Management Science,* **45**:1378–1398.

Kouvelis P., Zhao W. 2008. Financing the newsvendor: Supplier vs. bank, optimal rates, and alternative schemes. Working paper. St. Louis, Missouri: Washington University.

Lederer P.J., Singhal V.R. 1994. The effect of financing decisions on the choice of manufacturing technologies. *International Journal of Flexible Manufacturing Systems* **1**: 339–371.

Li L., Shubik M., Sobel M.J. 1997. Control of dividends, capital subscriptions, and physical inventories. Working paper. New Haven, Connecticut: Yale School of Management, Yale University. Available at: http://ssrn.com/abstract=690864.

Xu X., Birge J.R. 2004. Joint production and financing decisions: Modeling and analysis. Working paper. Evanston, Illinois: Northwestern University.

Xu X., Birge J.R. 2006. Equity valuation, production, and financial planning: A stochastic programming approach. *Naval Research Logistics,* **53**:641–655.

Xu Y., Zhang J. 2006. On the selection of supply chain coordinating contracts: The role of capital constraints. Working paper. Coral Gables, Florida: University of Miami.

Zhou J., Groenevelt H. 2007. Impacts of financial collaboration in a three-party supply chain. Working paper. Rochester, New York: The Simon School, University of Rochester.

PART FOUR

Operational Risk Management Strategies

CHAPTER FOURTEEN

Decentralized Supply Risk Management

GÖKER AYDIN, VOLODYMYR BABICH, DAMIAN BEIL, AND ZHIBIN YANG

14.1 Introduction

A paramount concern of today's supply chain managers is building supply chains that can handle supply disruptions. There are a number of reasons why supply chain managers are becoming increasingly preoccupied with supply risk. First, supply disruptions are more likely than before, because the widespread use of outsourcing is not only stretching supply chains further geographically, but it is also turning supply networks into intricate webs of highly interdependent players. In fact, in a 2008 survey of 138 companies, 58% reported that they suffered financial losses within the last year due to a supply disruption.[1] Second, outsourcing to external vendors is making supply risks harder to foresee and, therefore, harder to prepare for. Third, the consequences of supply risks have arguably become more costly than before. Successful initiatives such as lean manufacturing, quick

[1] For more details, see http://www.infoworld.com/article/08/09/17/Most_companies_lag_in_supply_chain_risk_management_1.html.

Handbook of Integrated Risk Management in Global Supply Chains, First Edition. Edited by Panos Kouvelis, Lingxiu Dong, Onur Boyabatli, and Rong Li.

response, and postponement proved beneficial in maintaining high fill rates while squeezing inventory out of the pipeline, but they also reduced the buffers that a firm could fall back on in the event of a supply disruption, exacerbating the costly effects of disruptions.

This chapter focuses on supply risk management in decentralized networks where self-interested firms are interacting. In this introduction section we first illustrate several types of risk. We then discuss the operational tools used to manage those risks. We emphasize the challenges and opportunities in supply risk management arising from the decentralized nature of the supply chain and highlight how supply risks influence the interactions among firms in supply networks. We review insights into decentralized supply risk management from the extant academic research and point out important future research directions.

14.1.1 TYPES OF SUPPLY RISK

The causes of supply disruptions are myriad, including accidents at supplier facilities, natural disasters, bankruptcy of a key supplier, defective parts or components, labor strikes, and so on. Despite the diversity of causes, supply risks generally fall into three categories depending on how they manifest themselves: shortage of a critical part or loss of supplier capacity loss of finished goods inventory due to the use of a defective part and inflated supply cost.

14.1.1.1 Shortage of a Critical Part or Loss of Supplier Capacity. Supply risk events often take the form of parts shortages or loss of supplier capacity. Such events arise in various settings and for a variety of reasons, as the following examples illustrate.

A well-documented instance of parts shortages is the experience of Boeing in the late 1990s. During that period, Boeing had trouble keeping up with demand for commercial aircraft and missed several delivery deadlines. The poor delivery performance was blamed mainly on shortages of parts such as tie rods and bearings (Biddle 1997b).

In some cases a loss of supplier capacity may occur due to a shift of the supplier's business strategy. A case in point is the medical device manufacturer Beckman Coulter's loss of its supplier Dovatron, who produced customized chips for Beckman Coulter.[2] After Dovatron was acquired by Flextronics in 2000, Flextronics restructured itself to focus on higher-volume products, and decided it would no longer serve Beckman Coulter, who was purchasing a low-volume specialty product.

There are abundant examples of supplier bankruptcies that threatened to cut the supply of critical parts. For example, bankrupt automotive supplier Collins & Aikman halted the shipment of parts to Ford's Fusion plant (McCracken 2006). In some cases, creditors of bankrupt suppliers may try to "hold up" manufacturers, taking advantage of the critical role many suppliers play for manufacturers. A

[2] For more details, see `http://www.callahan-law.com/verdicts-settlements/fraud-beckman-coulter/index.html`.

well-known example of a "hold-up" is UPF Thompson vs. Land Rover (Jennings 2002). In 2001, UPF Thompson, who was the sole supplier of chassis for Land Rover's Discover model, declared bankruptcy and was taken over by KPMG. KPMG then threatened to halt chassis shipments unless Land Rover made an additional $35 million payment to KPMG.

Likewise, there are many examples of accidents and natural disasters resulting in temporary loss of supplier facilities: Examples include the 1999 earthquake in Taiwan that disrupted semiconductor plants making 70% of the world's graphics chips and 10% of the world's memory chips (Savage 1999; Papadakis 2003); the 1997 fire at Aisin Seiki, the sole supplier of a key component used in the brake system of many Toyota models (Nishiguchi and Beaudet 1998); and the 2000 fire at a Phillips chip plant, which supplied both Nokia and Ericsson (Sheffi 2005).

A related but distinct type of supply risk is delivery lead-time risk. In Section 14.7, we discuss this type of risk, which is understudied in the literature.

14.1.1.2 Loss of Finished Goods Inventory due to the Use of a Defective Part. Another form of supply risk is the use of a defective input, which results in finished goods that do not meet the buyer's standards. Such supply risks can have very serious consequences and, if not caught early, can result in the recall of deployed finished goods inventory. For example, following the deaths of numerous pets in 2007, pet food producer Menu Foods Corp. had to recall more than 60 million cans and pouches of dog and cat food of more than 100 pet-food brands (Myers 2007). The deaths were later linked to melamine, a poisonous industrial chemical. The melamine was traced to wheat gluten, which Menu Foods (a Canadian firm) had bought from ChemNutra (a U.S.-based supplier), who, unbeknownst to Menu Foods, had outsourced it to Xuzhou Anying Biologic Technology Development Co. Ltd. (a Chinese supplier). In the past few years there have been many other similar recalls: Examples include the spinach recall in the U.S. spurred by the discovery of batches contaminated with E.coli (Wall Street Journal 2007), Mattel's recall of toys covered in lead paint (Casey 2007), and several computer manufacturers recalling their laptops due to defective batteries produced by Sony (Morse 2006).

14.1.1.3 Inflated Supply Costs. Another type of supply risk is related to uncertainty about the cost of inputs. For example, a part that is procured from a distant supplier may quickly become more expensive if rising oil prices lead to a hike in shipping costs. Supply cost risks can also arise in conjunction with shortages. For instance, in 2000, the price of palladium increased sharply when Russia, the main source of this precious metal—held up its supply. Consequently, automotive manufacturers, who use palladium in catalytic converters, suffered a $100 increase in per-vehicle production costs (White 2000).

14.1.2 MANAGING SUPPLY RISK

Firms use various tools to manage supply disruption risks. These tools may be proactive (e.g., supplier qualification screening/risk discovery, diversification, or investing in suppliers to reduce the odds of a supply disruption) or reactive (e.g., using external or internal backup sources of supply in the face of a disruption, or levying nonperformance penalties on suppliers). We next discuss each of these operational risk management tools.

14.1.2.1 Supplier Qualification Screening/Risk Discovery. A buyer can perform qualification screening when selecting a new supplier or can audit its existing suppliers. Such measures enable the buyer to learn about the supplier, thus allowing the buyer to identify and avoid or rectify weaknesses that may potentially cause disruptions. For example, had Menu Foods or Mattel (see examples above) done thorough and frequent supplier audits, they might have caught the problems that later resulted in recalls.

14.1.2.2 Investing in Suppliers to Reduce the Odds of a Disruption. Buyer investments in suppliers can take many forms, ranging from helping a supplier improve its production processes to providing funds to a financially unstable supplier to avoid imminent bankruptcy. The latter is what Ford chose to do when faced with the danger that its supplier, Visteon, would go bankrupt: Ford agreed to pay up to $1.8B to avert a bankruptcy (White 2005).

14.1.2.3 Multisourcing. In many cases a buyer will have the option to source the same part from not one, but multiple suppliers. When such an option is exercised (i.e., when the buyer diversifies), the buyer is less vulnerable to risks associated with any one supplier. As such, diversification can help make the buyer more resilient to supply risks such as shortages, defective parts, or loss of supplier capacity. On the other hand, when the buyer chooses not to diversify, multisourcing results in increased supplier competition, thus yielding benefits to the buyer.

14.1.2.4 Using External or Internal Backup Sources of Supply. In the face of a disruption, the buyer can scramble to create an alternate source of supply. For example, when the fire at Aisin Seiki threatened to halt the production of many Toyota models, Toyota and Aisin Seiki worked together with many other suppliers to create an alternate source. Likewise, when Beckman Coulter lost its supply of chips from Dovatron, it chose to replace the lost supply by building its own in-house production line.

14.1.2.5 Nonperformance Penalties (or Payments Contingent on Supply Events). Most supply contracts include provisions for penalties or nonpayment that will be imposed on a supplier in the event that the supplier fails to deliver on its promises. As a last resort, a buyer can choose to sue the supplier to enforce such penalty clauses. Alternatively, the buyer may make payments to the supplier contingent on the product passing product inspection or defer paying supplier for some time to give the customers an opportunity to discover defects over time. An example of a deferred payment contract is a popular financial contract called *trade credit*.

14.1.3 THE ROLE OF DECENTRALIZATION

Most supply chains are networks of several self-interested firms. The decentralized nature of supply chains is an important consideration in managing supply risk, and it manifests itself in four important ways: misalignment of incentives between buyers and suppliers, competition among suppliers, competition among buyers, and asymmetric information.

14.1.3.1 Misalignment of Incentives Between Suppliers and Buyers. In many buyer–supplier relationships, the supplier's priorities and interests are not necessarily well aligned with those of the buyer. Such misalignments may cause or exacerbate supply risks. For example, the parts shortage that resulted in Boeing's troubles was partly due to suppliers not keeping up with Boeing's major overhaul of its production process to improve its cycle times (Biddle 1997a). Likewise, in the case of Beckman Coulter, the part that was highly critical to Beckman Coulter was simply an unprofitable specialty product for Dovatron, which is why Flextronics dropped the part after acquiring Dovatron.

14.1.3.2 Competition among Suppliers. Competition among suppliers may decrease as a result of efforts to prevent disruptions. For example, with a multisourcing option, when the buyer commits to ordering from only one supplier the suppliers compete harder to win the buyer's business, but when the buyer commits to diversification (i.e. ordering from multiple suppliers), the suppliers have less incentive to compete aggressively. As a general strategy to encourage competition, the buyer may wish to order from suppliers that are very similar, in particular, suppliers that are exposed to the same risks. However, this reduces diversification benefits. Furthermore, supplier competition can be an aggravating factor in causing disruptions. For example, cost competition among suppliers may result in suppliers cutting corners, thereby increasing the chances of defective parts. Similarly, fierce price competition among suppliers may get to the point where the suppliers' profitability is threatened so much as to cause financial instability in the supply base.

14.1.3.3 Competition Among Buyers. Competition among buyers in a decentralized system can present a number of challenges and opportunities. When multiple buyers rely on the same supplier, competition among buyers may exacerbate the ramifications of a supply disruption. A case in point is the 1999 earthquake in Taiwan, which resulted in a major shortage of chips. This shortage increased the competition for chips among PC manufacturers, causing a five-fold increase in the spot price of chips, thus inflating the input costs of major PC producers by as much as 25% (Papadakis 2003). On the other hand, a supplier working with multiple buyers could be more financially viable because it receives multiple subsidies from these buyers. While all buyers may have a stake in keeping the supplier solvent, some might be reluctant to do so, because by helping the joint supplier they are also helping their competitors. Buyers' willingness to support the supplier depends on the volume of business they have with the supplier (e.g., automotive manufacturers may all utilize the same supplier, but to different extents).

14.1.3.4 Asymmetric Information. In a decentralized supply chain, information about a supplier's actions of vulnerability to a disruption is not readily available to a buyer. Not having access to this information may be very detrimental to a buyer. For example, Land Rover was not aware of the looming bankruptcy of its supplier UPF-Thompson. Hence, once the bankruptcy happened, Land Rover was unprepared and had few options (Jennings 2002). Menu Foods was not aware that its first-tier supplier, ChemNutra, outsourced to a second-tier supplier. It was this second-tier supplier that adulterated the product "to stretch the supply" and introduced melamine into the product in an effort to hide the adulteration. The melamine was to be the cause of the pet food contamination (Myers 2007). Unfortunately, suppliers often will not (or cannot) voluntarily (or credibly) share their reliability information, and buyers have to work to elicit it.

The remainder of this chapter is organized as follows. Section 14.2 overviews literature related to decentralized supply risk management. Subsequent sections address key issues that arise in managing supply risk in decentralized supply chains. Section 14.3 discusses misalignment of incentives, Section 14.4 addresses supplier competition, Section 14.5 addresses buyer competition, and Section 14.6 addresses asymmetric information. A discussion of possible future work in decentralized supply risk management is provided in Section 14.7.

14.2 Literature Taxonomies

There are a number of excellent reviews of the general supply-risk management literature. The most recent is Tomlin and Wang (2010). In this chapter we focus on the subset of the supply-risk management literature that deals with decentralized systems. Insights obtained from this work will be discussed in more detail in this chapter's subsequent sections. In the present section we provide a taxonomy of the decentralized supply risk management literature. This taxonomy is presented in Table 14.1, which encapsulates the coming discussion in this section.

14.2.1 DECENTRALIZATION TYPES

One classification dimension for the literature taxonomy is the type of decentralization modeled. As discussed in Section 14.7, decentralization manifests itself in supply networks in a number of ways: misalignment of incentives between buyers and suppliers, competition among suppliers, competition among buyers, and asymmetric information. The first of these—incentive misalignment—is present in virtually every paper about decentralized supply risk management (see Section 14.3). However, the other three aspects of decentralization appear separately.

TABLE 14.1 A Summary of the Literature on Decentralized Supply Risk Management

	Decentralization Effect				Operational Tool			
Papers	**Asymm. Information about Supplier Reliability**	**Supplier Competition**	**Buyer Competition**	**Supplier Qualification Screening/risk Discovery**	**Investing in supplier Reliability**	**Multi-sourcing**	**Creating Backup Sources of Supply**	**Non-Performance Penalties**
Babich (2006)		×				×	×	
Babich et al. (2007)		×				×		
Babich and Tang (2010)	×							×
Baiman et al. (2000)					×			×
Chaturvedi and Martínezde-Albéniz (2010)	×	×				×		×
Deo and Corbett (2008)			×					
Federgruen and Yang (2009)		×			×	×		
Gurnani and Shi (2006)	×							×
Lim (1997)	×							
Swinney and Netessine (2009)					×			
Tang and Kouvelis (2009)			×			×		
Wadecki et al. (2010)			×		×			
Wan and Beil (2008)	×	×		×				
Wan and Beil (2009)	×	×		×				
Yang et al. (2009a)	×			×			×	×
Yang et al. (2009b)	×	×		×		×		×
Zimmer (2002)					×			×

14.2.1.1 Competition Among Suppliers. A subset of the literature on supply-risk management explicitly models the presence of two or more suppliers and the competitive interactions among them. This stream of research includes Babich et al. (2007), Babich (2006), Chaturvedi and Martínez-de-Albéniz (2010), Federgruen and Yang (2009), Wan and Beil (2008, 2009), and Yang et al. (2009b). For discussion of supplier competition models, please see Section 14.4.

14.2.1.2 Competition Among Buyers. An emerging stream of research recognizes that multiple buyers may compete for uncertain supply (e.g., Deo and Corbett (2008), Tang and Kouvelis (2009), Wadecki et al. (2010). See Section 14.5 for detailed discussion of models incorporating buyer competition.

14.2.1.3 Asymmetric Information. Different types of asymmetric information relevant to supply risk have been studied. Yang et al. (2009a), Chaturvedi and Martínez-de-Albéniz (2010), and Yang et al. (2009b) study asymmetric information about the supplier's likelihood of experiencing a disruption and about the supplier's production cost. Baiman et al. (2000) and Lim (1997) study asymmetric information about the supplier's ability to produce nondefective items. Wan and Beil (2009) consider asymmetric information about suppliers' production cost. Babich and Tang (2010) study asymmetric information about a supplier's actions to adulterate products. A detailed description of asymmetric information and supply risk models is provided in Section 14.6.

14.2.2 TYPES OF RISK MANAGEMENT TOOLS

Another literature classification dimension is the type of risk-management tool studied. Section 14.1.2 listed a number of tools that can be used to manage supply risks: supplier qualification screening/risk discovery, multisourcing, investing in suppliers to reduce the odds of a supply disruption, creating external or internal backup sources of supply in the face of a disruption, and using nonperformance penalties. The following is a classification of papers that study these tools.

14.2.2.1 Supplier Qualification Screening/Risk Discovery. Wan and Beil (2009) model a setting where suppliers' qualifications are learned through a qualification screening process while Yang et al. (2009a,b) study models in which the buyer designs a menu of contracts to elicit the supplier's true reliability information. See Section 14.6 for a review of these papers.

14.2.2.2 Investing in Suppliers to Reduce the Odds of a Disruption. Babich (2010) and Wadecki et al. (2010) study models where the buyer provides financial subsidies to a supplier, who can then use subsidies to pay for its production costs. Swinney and Netessine (2009) allow a buyer to reduce the odds of a supplier bankruptcy by offering a long-term contract.

14.2.2.3 Multisourcing. Examples of papers that study multisourcing are Babich et al. (2007), Babich (2006), Chaturvedi and Martínez-de-Albéniz (2010), Federgruen and Yang (2009), Tang and Kouvelis (2009), Wan and Beil (2008), and Yang et al. (2009b). These papers are reviewed in Section 14.4.

14.2.2.4 Creating External or Internal Backup Sources of Supply. Babich (2006) and Yang et al. (2009a) allow the possibility that either the buyer or the supplier or both have access to a source of backup production in the event of a disruption.

14.2.2.5 Nonperformance Penalties (and Contingent Payments). Zimmer (2002), Chaturvedi and Martínez-de-Albéniz (2010), and Yang et al. (2009a,b) investigate the use of penalties imposed on a supplier in the event of nondelivery. Babich and Tang (2010) investigate the use of payments contingent on favorable inspection outcomes or customers not discovering product defects over a prespecified period in inspection, deferred payment, and combined mechanisms.

14.2.3 TYPE OF SUPPLY UNCERTAINTY

Another classification dimension for the literature is the type of supply uncertainty. We listed several types of supply risks in Section 14.1.1: shortage of a critical part, loss of supplier capacity, loss of finished goods inventory due to a defect, and inflated supply cost. Some models in the literature (e.g., Deo and Corbett 2008, Federgruen and Yang 2009, Tang and Kouvelis 2009) use proportional random yield models, whereby if the buyer orders $\overline{q}$ units, it receives only a fraction $q = y\overline{q}$ of the quantity ordered, where y is a random number. Other papers (e.g., Zimmer 2002, Babich 2010) model supply risk as random capacity, in which case the supplier's random capacity may impede its ability to meet the buyer's order in full. That is, if the buyer orders $\overline{q}$ units, it receives $q = \min(\overline{q}, ky)$, where k is the supplier's regular capacity and y is a random shock. Usually random capacity models are more analytically tractable than random yield models. (For example, optimal dynamic policies with random yield uncertainty can be extremely complex, as discussed in Yano and Lee 1995, but with random capacity the standard order-up-to policy structure is optimal, as in Ciarrallo et al. 1994, Babich 2010). A popular subclass of yield/capacity models (e.g., Baiman et al. 2000, Babich 2006, Babich et al. 2007, Chaturvedi and Martínez-de-Albéniz 2010, Gurnani and Shi 2006, Swinney and Netessine 2009, Yang et al. 2009a,b) is that with all-or-nothing yields, which corresponds to random number y following a Bernoulli distribution, the buyer receives either the full order quantity or nothing at all. This type of yield model offers a fine compromise between tractability and fidelity, and is especially appropriate when modeling, for example, the loss of finished goods inventory due to recalls, or accidents or closures that put a supplier out of commission. Finally, supply cost risks have also been modeled in the study of decentralized supply risk management (Wan and Beil 2008).

14.3 Misalignment of Incentives

Every paper we review in this chapter features misalignment of incentives between the buyers and suppliers. Thus, in subsequent sections misalignment of incentives plays a key role, but for expositional purposes we organize these sections around other ways in which decentralization affects supply risk management, namely, supplier competition, buyer competition and asymmetric information. However, even without those effects, misalignment of incentives still plays an important role in decentralized supply chains. For example, Zimmer (2002) studies a problem where the supplier might not invest in as much capacity as the buyer would like it to. In the paper's model, only a random fraction of the supplier's regular production capacity will be available when the buyer places an order, but the supplier can proactively procure a reliable premium capacity to increase its ability to meet the buyer's requirement. The author provides the insight that the buyer can coordinate the supply chain using a contract with either a penalty term for shortage or a bonus term for on-time delivery.

14.4 Competing Suppliers

Supplier competition is possible when the buyer has several suppliers for the same product in its supply base, what we call "multisourcing." Competition among suppliers is good for the buyer because suppliers, keen on winning the buyer's business, are willing to accept smaller payments for their goods and services. Another benefit of multisourcing is risk reduction through diversification (i.e., ordering from multiple suppliers simultaneously). Diversification and competition are two primary reasons for multisourcing in practice (see Wu and Choi 2005, and references therein). An important question that managers should consider is how their decisions affect both diversification and competition benefits. Intuitively, competition is most intense when the suppliers know that only one of them will be awarded the buyer's business, but in this case the buyer does not get any diversification benefits. Conversely, if the buyer orders from multiple suppliers for the sake of diversification, the suppliers have less incentive to compete. Is it possible to enjoy both types of benefits or do managers have to sacrifice one benefit for the other one? As we will see, the answer furnished by academic research is: "It depends." To understand this answer, in the following we will define what we mean by supplier competition, discuss various forms and models of competition, illustrate the benefits of diversification, and look at how competition and diversification interact with each other under different competition models.

14.4.1 FORMS AND MODELS OF SUPPLIER COMPETITION

Supplier competition describes interactions among suppliers at the same horizontal level of the supply chain, with the property that the suppliers' interests are conflicting and the actions of one supplier can affect other suppliers' payoffs.

Supplier competition can take numerous forms. Babich et al. (2007) and Babich (2006) use a Bertrand competition model (Bertrand 1883), where risky suppliers submit bids for the per-unit wholesale price of the product and the buyer selects which suppliers to work with and how much to order from each supplier, based on these bids and the joint distribution of supplier defaults. Babich et al. (2007) consider a one-period model, whereas Babich (2006) uses a multiperiod model and recognizes that disparities in production lead times across suppliers allows the buyer to delay its ordering decision. More specifically, in the latter paper, the buyer has a deferment option, which is "vulnerable," because the supplier with shorter lead time may experience a default before the buyer has a chance to order from it. Babich (2006) considers a multistage game, where the "slow" supplier submits a bid, then the buyer makes an ordering decision, then the "fast" supplier submits a bid, then the buyer makes another ordering decision. Thus, supplier competition is separated in time, which leads to interesting observations, as will be discussed below.

Instead of competing on price while keeping the supply risk parameters constant (as in Babich et al. [2007] and Babich [2006]), suppliers can compete on "quality" by making investments in their production capabilities that affect their yield uncertainty. Federgruen and Yang (2009) study examples of quality competition, in which the supplier selects the mean and/or standard deviation of their production yield distribution. The buyer must satisfy a service level constraint (in contrast to the majority of other papers where shortage penalties are captured in the buyer's objective function) and the buyer's ordering decisions are determined by applying a Normal approximation to the total quantity received from the suppliers. Federgruen and Yang (2009) point out that these examples of supplier competition are special cases of generalized attraction models of competition, which appear in a number of other contexts, such as advertisement competition (Karnani 1985) and fill-rate competition.

Suppliers need not make decisions beyond which contract to choose from a menu offered by the manufacturer in order for competition effects to be present. This kind of competition effect is highlighted in recent papers studying buyer-supplier relationships where the supplier has private information about its costs and supply risks. In Yang et al. (2009b) and Chaturvedi and Martínez-de-Albéniz (2010), the buyer does not know the suppliers' types. Suppliers' types determine their production reliabilities and production costs. The buyer designs a menu of contracts to offer to the suppliers. In Chaturvedi and Martínez-de-Albéniz (2010) the supplier simply decides which of the buyer's contracts to accept (if any). In Yang et al. (2009b) the supplier decides which of the contracts to accept (if any) and then proceeds to make production decisions to fulfill the contract. If the buyer knew the suppliers' costs and reliabilities, it could extract the entire supply chain profit, leaving all suppliers with zero profits. But under asymmetric information, the suppliers can misrepresent their cost or reliability, permitting them to earn positive profit (see Section 14.6). Competition between suppliers in a two-supplier model arises when the buyer promises to select only one of the suppliers to fulfill the order, curbing suppliers' misrepresentation incentives and reducing the suppliers' profits. Thus, the presence of other suppliers affects profits in a way that is akin to the suppliers bidding against each other for the buyer's business.

One way to implement the bidding process is through auctions. Wan and Beil (2008, 2009) use auctions as a way of allocating orders to suppliers. For example, in Wan and Beil (2008), suppliers privately know their own production cost. Furthermore, there is also a publicly known region-specific production cost, which depends on the geographic location of the supplier (e.g., U.S. vs. Mexico vs. China). The buyer first designs its supply base (decides from which regions it would like to qualify suppliers) and then in each period, once the uncertainty about current region-specific costs have been realized, runs an auction to decide to whom to award the business.

What do we learn from these various competition models about managing supply risk? To understand the role of competition, it helps to understand the diversification benefits without competition first.

14.4.2 DIVERSIFICATION BENEFITS

In this subsection we will illustrate the benefits of risk reduction through diversification. A buyer can make its supply more certain by ordering from two or more suppliers instead of one, provided the causes of supply disruptions are not perfectly correlated (recall our discussion in Section 14.1 on causes of disruptions). If the buyer's revenue is concave (exhibits a diminishing rate of returns) with respect to the supply quantity, or if its cost is convex (its marginal cost is increasing), the buyer favors a more certain supply quantity, as shown in the following example. (A side note: generally speaking, for diversification and risk management to have value, one does not need to assume that decision makers are risk-averse; as long as the objective function of the maximizer is concave or the objective function of the minimizer is convex in random variables of the problem, risk management is valuable; for more discussion on this see Froot et al. [1993].)

EXAMPLE 14.1

Consider a buyer facing a demand of one unit for the next period. Assume the per-unit revenue for the product is \$100. There are two potential suppliers of the product with production costs of \$20 for supplier 1, and \$21 for supplier 2, and disruption probabilities of 0.4 for both suppliers. If a disruption happens, the entire order is lost, but if no disruption happens the buyer receives the full order. The buyer has to pay for the product up front, but it has the power to set wholesale prices (and in particular it can set wholesale prices to be equal to the suppliers' production costs). The suppliers will accept the contract from the buyer as long as wholesale prices exceed or equal their respective production costs. Thus, in the optimal (for the buyer) contract the wholesale prices will be equal to the suppliers' production costs.

To highlight the benefits of diversification, first assume that supplier disruptions are perfectly correlated. That is, if one supplier is down, the other one is down as well, and if one supplier is up, the other one is up as well. The buyer has no benefit from ordering from two suppliers and it will decide with which supplier to work by comparing the expected profits when working with each supplier. Of these two suppliers the buyer would choose to work with supplier 1, because it is cheaper, and the buyer's expected profit by ordering z units is:

$$\text{Expected revenue} - \text{order cost} = \\ 0.6\,\$100\min(1, z) + 0.4\,\$100\min(1, 0) - \$20z$$

The optimal choice for the order quantity is $z = 1$, leading to an expected profit of \$40.

In contrast, suppose that supplier disruptions are independent (this implies that the probability of both suppliers experiencing disruptions is $0.4 \cdot 0.4 = 0.16$, the probability that a given supplier experiences a disruption and the other one does not is $0.4 \cdot 0.6 = 0.24$, and the probability that both suppliers deliver is $0.6 \cdot 0.6 = 0.36$). If the buyer orders from one supplier, then the expected profits remain the same as before (\$40 with supplier 1 and \$39 with supplier 2). However, if the buyer orders 1 unit from each supplier, it hedges its risk that one of the suppliers will not deliver and earns expected profit:

$$\text{Expected revenue} - \text{order cost} = \\ \$100[0.36\min(1, 2) + 0.24\min(1, 1) + 0.24\min(1, 1) \\ + 0.16\min(1, 0)] - \$20 \cdot 1 - \$21 \cdot 1 = \$43$$

Thus, it is optimal for the buyer to order 1 unit from each supplier. The extra \$3 (= \$43 − \$40) of expected profit when ordering from both suppliers (relative to the best profit when ordering from one supplier only) is the benefit of diversification. This benefit decreases in the correlation between supplier disruptions (e.g., this benefit is zero when supplier disruptions are perfectly correlated), in accordance with intuition about diversification benefits in financial portfolios.

Having illustrated diversification benefits, we will now discuss how competition and diversification interact.

14.4.3 IS POSITIVE CORRELATION BETWEEN DISRUPTIONS BAD FOR THE BUYER?

The first important insight is that competition and diversification can work against each other. The following simple illustrative example is based on Babich et al. (2007):

EXAMPLE 14.2

Use the same data as in Example 14.1, but assume that the suppliers competitively choose wholesale prices.

When supplier disruptions are perfectly correlated, the buyer will never choose to work with more than one supplier. Instead, the buyer will order from the supplier who offers the lowest wholesale price. Thus, suppliers effectively compete in price for a single order from the buyer (this is the classical Bertrand competition model). Supplier 1 will win the order by offering a wholesale price just below \$21, a price impossible for Supplier 2 to profitably match. Supplier 1 will earn profit just short of \$1, supplier 2 will be priced out and earn zero profit, the buyer will order from supplier 1 and will earn expected profit just above:

$$\text{Expected revenue} - \text{order cost} =$$
$$0.6\,\$100\min(1,1) + 0.4\,\$100\min(1,0) - \$21 \cdot 1 = \$39$$

Note that, compared with Example 14.1, the buyer's profit is \$1 lower. However, \$39 is still much higher than \$0 profit the buyer would earn if supplier 1 were a monopolist. The reason the buyer can keep most of its profit from Example 14.1 is competition between suppliers, which keeps the wholesale prices at \$21.

Next, suppose that supplier disruptions are independent. Figure 14.1 illustrates the buyer's optimal decision as a function of the wholesale prices

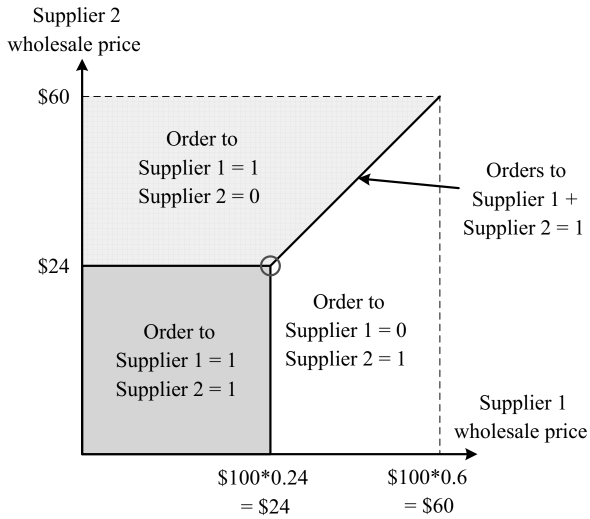


FIGURE 14.1 Buyer's order decisions as a function of supplier 1 and supplier 2 prices. The origin of this figure is at the supplier's unit costs, 20 and 21 for suppliers 1 and 2, respectively.

the suppliers bid. One can show (see Babich et. al. 2007) that the buyer is indifferent between diversifying and not diversifying if both suppliers offer wholesale prices $\$100 \cdot 0.24 = \$24$ (recall that the probability of a given supplier delivering and the other one failing is 0.24). Let's say the buyer would diversify, if indifferent. If both suppliers bid below \$24, the buyer will order from both suppliers (bottom left square in the figure). If supplier 2 bids higher than \$24 and supplier 1 bids less than supplier 2's bid, then only supplier 1 will receive an order (upper left part of the figure). Similarly, if supplier 1 bids higher than \$24 and supplier 2 underbids supplier 1, only supplier 2 will receive an order. No supplier who bids above \$60 will receive an order.

We claim that \$24 is the equilibrium wholesale price in this simple economy. If one of the suppliers lowers its price from \$24 to, say, \$23, that supplier will still receive the same order for one unit from the buyer, but will earn \$1 less in profit. Thus, no supplier would want to lower its price unilaterally. If one of the suppliers raises its price from \$24 to, say, \$25, the buyer will no longer order from that supplier, thus no supplier will want to raise its price unilaterally. With equilibrium wholesale prices of \$24, the profit of supplier 1 is $(\$24 - \$20) \cdot 1 = \$4$, the profit of supplier 2 is $(\$24 - \$21) \cdot 1 = \$3$, and the expected profit of the manufacturer is

$$\begin{aligned}&\text{Expected revenue} - \text{order cost} =\\ &\qquad \$100[0.36 \min(1, 2) + 0.24 \min(1, 1) + 0.24 \min(1, 1)\\ &\qquad\qquad + 0.16 \min(1, 0)] - \$24 \cdot 1 - \$24 \cdot 1 = \$36.\end{aligned}$$

First, observe that as the correlation between defaults decreased (from perfect to zero), contrary to what we have seen in Example 14.1, the buyer's expected profit decreased and the suppliers' profits increased. Second, the key to understanding why this has happened lies in the increase of the equilibrium wholesale prices from \$21 to \$24. This increase in prices is a manifestation of reduced competition among suppliers, which does not exist in Example 14.1 or in other traditional "portfolio optimization" models.

An intuitive way to understand this example is to realize that the suppliers offer a product, which is identical in every respect, except for the states of nature in which this product will be delivered for the buyer. The correlation between defaults can be interpreted as the overlap between the states of nature where both suppliers deliver. If the supplier defaults are perfectly positively correlated, then the products from the two suppliers are perfect substitutes (the products will be available or not available in the same states of nature). In this case suppliers are engaged in classical Bertrand competition and the supplier with the lowest cost will charge a wholesale price equal to the second-lowest cost. Although there is no diversification benefit, the buyer enjoys low wholesale prices. If the supplier defaults are not perfectly correlated, the buyer enjoys some diversification benefits, but because the products are no longer perfect substitutes, the prices the suppliers charge are higher.

The idea that the states of nature where a supplier delivers are a part of the product's attributes is an important takeaway and it will be used again in our discussion of buyer competition in Section 14.5.

14.4.4 BUYER'S POWER TO DESIGN CONTRACTS AND COMMITMENT TO COMPETITION

In Example 14.2 the buyer plays a somewhat passive role. It does not control the mechanism by which prices are set and orders are allocated (i.e., the buyer does not design an auction or order allocation mechanism). The buyer decides its order quantities after receiving the suppliers' price bids, but does not have the power to commit a priori (before seeing the prices) to ordering from only one supplier. In Yang et al. (2009b) the buyer plays a more active role and, by designing the allocation mechanism (the rules governing the order quantities and fixed and variable payments it offers), may commit to sole sourcing. This power to commit proves beneficial for the buyer. The buyer may find it optimal to forego diversification in order to get better pricing from suppliers, who know they must compete harder to secure an order from the buyer. Thus, a key insight is that in a supply risk environment, where diversification and competition work against each other, the buyer may find it optimal to sole source more in order to foster competition.

When the buyer foregoes diversification in order to enjoy the benefits of competition, consumers and the supply chain (as a whole) are worse off: Without diversification, the quantity that the supply chain provides to consumers is reduced (where reduction should be understood in the stochastic ordering sense, as this quantity is a random variable).

In Example 14.2 and Babich et al. (2007), correlation was a key driver of supplier competition and the buyer could not control the contract between it and the suppliers. In Yang et al. (2009b), the buyer can design the optimal mechanism for its interactions with the suppliers and correlation across suppliers' disruption risks does not increase competition, only the buyer's decision to sole source does (but the buyer is more likely to sole source as the correlation between supplier disruptions increases). Therefore, unlike Babich et al. (2007), in Yang et al. (2009b) the buyer prefers less correlated supplier disruptions, in order to make diversification—when it is used—more valuable.

The broader idea that the ability to control procurement mechanisms might affect the buyer's preference for competition, diversification, and risk correlation was first studied in Wan and Beil (2008) in the framework of uncertainty about supply cost risks. Thus far we have discussed uncertain supply availability, where diversifying meant ordering from multiple suppliers. Another possibility is that supply availability is certain, but the buyer is concerned about cost risks. Even if the buyer commits to sole sourcing, having suppliers in different regions bidding for the buyer's business can "diversify" the supply base. The following example, based on Wan and Beil (2008), illustrates the trade-off between competition and diversification benefits in such an environment. As in Wan and Beil (2008), the example reveals how this trade-off is affected by the buyer's ability to choose the contract allocation mechanism.

Example 14.3

Consider a buyer who will run a total-cost reverse-English auction (see explanation below) to award a contract to one of two suppliers. The buyer does not know the suppliers' true production costs x_1 and x_2, but does know how much it would cost the buyer to transport the goods from either supplier.

The total-cost reverse-English auction runs as follows: Suppliers take turns bidding prices, the buyer adds on the transportation costs, and bidders can view the resulting total cost bids in real time. The auction ends when no bidder is willing to lower their price bid further, and the lowest total-cost bidder is the winner and is paid its final price bid.

Suppose the buyer is located in Europe, and both suppliers are located in Taiwan (so the transportation cost from either supplier is the same, say \$4 per unit). Because transportation costs are the same for both suppliers, production cost alone determines the auction winner. Assume without loss of generality that $x_1 = \$8 < \$10 = x_2$ (prices are per-unit); then supplier 1 wins at supplier 2's dropout bid, \$10. The buyer's total per-unit cost (payment to winner plus transportation cost) is \$10 + \$4 = \$14.

Transportation costs differ from region to region and are subject to shocks.[3] Intuitively, choosing suppliers in different regions mitigates the buyer's exposure to the risk of high transportation costs. To see how a more diverse supply base would affect the buyer's total costs, assume the same numbers as before, except suppose supplier 1 is located in Brazil from which the transportation cost is \$1 per unit. Now supplier 1's total (production plus transportation) cost is \$8 + \$1 = \$9, while supplier 2's total cost is still \$10 + \$4 = \$14. Supplier 2 will drop out of the auction with a price bid of \$10, translating to a total cost bid of \$14. Supplier 1 will defeat supplier 2 by offering a price just below \$13. The buyer's total per-unit cost (payment to winner plus transportation cost) is \$13 + \$1 = \$14.

Note that, despite the diversified supply base, the buyer incurs the same total cost as before. Although the buyer enjoys cheaper transportation costs (\$1 instead of \$4), this is offset by supplier 1 charging a higher per-unit price (\$13 instead of \$10). When the buyer must rely solely on supplier competition for price concessions, diversifying the supply base (having suppliers in different regions) does not necessarily help the buyer. In fact, one can show that it actually makes the buyer worse off by increasing the buyer's

[3] For example, in 2008 insurance premiums on Asia to Europe ocean transport increased tenfold, as ships passing through the Suez Canal faced piracy threats off the Somali coast (Costello 2008).

cost risks: Whenever either region experiences a high transportation cost, the supplier in the other region can simply raise its price and enjoy high windfall profits at the expense of the buyer.

Again, we see that the buyer may prefer less diversification (more correlated supplier costs) in order to foster more competition. The key feature for Example 14.3 is that the buyer must rely solely on supplier competition for price concessions. As Wan and Beil (2008) explain, if the buyer can institute other techniques to limit suppliers' windfall profit-taking (such as optimal auction rules admitting only part of a supplier's regional cost advantage over its competitors), then the buyer can retain much more of the diversification benefits that arise from having suppliers in different regions, and such buyers prefer to diversify (prefer less correlated supplier costs).

Thus, a key insight is that buyers with greater power to design contract allocation mechanisms find diversification more attractive. To put this insight in the context of supply availability risks, note that in Example 14.1 the buyer had the power to make take-it-or-leave-it offers and preferred to diversify when the suppliers' disruption risks were independent. In contrast, in Example 14.2 the buyer relied on competition for price concessions, and would have preferred to commit to sole sourcing (had it been possible to do so).

14.4.5 CAN A BUYER HAVE BOTH COMPETITION AND DIVERSIFICATION BENEFITS?

We see from Examples 14.2 and 14.3 that the buyer might prefer suppliers that have more correlated risks (on either costs or supply disruptions). While seemingly counter intuitive, this insight is based on the fact that more correlated suppliers compete harder with each other and are less able to use high pricing to devour the benefits of diversification. An interesting corollary is that the buyer might be able to enjoy both the benefits of competition and diversification (as Babich et al. 2007 and Wan and Beil 2008 show) if the buyer has, say, four suppliers and "partially diversifies." That is, divide suppliers into clusters (e.g., geographically) of two or more suppliers each. Within each cluster the disruptions must be highly correlated (e.g., all Taiwanese LCD suppliers are exposed to the same natural disaster, political instability, country-wide financial crisis risks), so that there is a fierce intercluster competition among suppliers. At the same time, correlation of disruptions across clusters must be low, so that by ordering from suppliers from different clusters (e.g., from Taiwan and the U.S.), the buyer can enjoy the benefits of diversification.

14.4.6 WEAKENING SUPPLIER COMPETITION BY SEPARATING DECISIONS IN TIME

Competition among suppliers depends on products they offer (i.e., the degree of product substitutability), the correlation among suppliers (i.e., supplier disruptions can be thought of as a product attribute), as well as the timing of the bids. In the context of supply disruptions, suppliers can weaken competition if their bids and the buyer's orders are spread over time, which is another managerial takeaway. Recall that Babich (2006) considers a multistage game between the buyer and two suppliers: one supplier with a short production lead time and the other supplier with a long production lead time. In this game, the "slow" supplier submits a bid, then the buyer makes an ordering decision, followed by the "fast" supplier submitting a bid, and the buyer making another ordering decision. When suppliers have the power to set wholesale prices, this order deferment by the buyer weakens competition between suppliers, to their advantage. Unlike results in Babich et al. (2007) discussed above, in this paper the buyer no longer prefers perfect correlation between supplier disruptions. Instead, it earns the highest profit at intermediate correlation values. Another interesting observation arising from order deferment is that, when the buyer holds all of the bargaining power (sets wholesale prices), a supplier might suffer if it reduces its lead time. Specifically, under certain conditions, if both suppliers have the same lead time, they would both receive orders from the buyer. But if one of the suppliers reduces its lead time the buyer would treat that supplier as a backup and order from it only if the slower supplier had a disruption.

14.4.7 MULTIPLICITY OF EQUILIBRIA

One consideration to keep in mind when interpreting the results from game theoretic models is that there could be multiple equilibria. This observation is important for managers because (1) using game theoretic models for decision making is difficult if one does not know which of the equilibria will be realized in practicea and (2) benefits to the system could be significant if there is a way of affecting which equilibria is realized. Illustrating this point in the context of supplier competition, Federgruen and Yang (2009) derive multiple equilibria in their competition models. Furthermore, they show that different equilibria could be quite disparate from the perspective of social welfare. However, their game model is log-supermodular, which allows them to derive comparative statics for some of the equilibria. They show that if some quality standards can be imposed by the government or other outside entity, the highly undesirable equilibria can be eliminated.

14.4.8 SECTION SUMMARY

The key insights and main takeaways from this section are as follows. First, as academic researchers and practitioners know, there are benefits of diversifying the supply base due to supply risk reduction and these benefits increase as the correlation among random events decreases. Next, decreasing correlation can be thought of as decreasing substitutability among products. Therefore, if the suppliers compete, as correlation decreases, the competition becomes less intense. Consequently, buyers may actually prefer suppliers with highly correlated disruptions, because such suppliers compete the most. A buyer who commits to sole sourcing encourages competition, but abandons diversification. In general, the more control the buyer has over the design of the mechanism that governs its interactions with suppliers, the less the buyer has to rely on competition among suppliers to keep its procurements cost in check and the more the buyer enjoys diversification benefits. When working with multiple suppliers the buyer may be able to capture both competition and diversification benefits by breaking suppliers into clusters with fierce inter-cluster competition and diversification across clusters. Ordering from multiple suppliers over time can be advantageous to the buyer because this provides flexibility to respond to demand and supply fluctuations. At the same time the competition among suppliers weakens if their orders are not placed simultaneously. Finally, a modeling takeaway is that competition models may have multiple equilibria as their solutions and properties of these equilibria can vary.

14.5 Competing Manufacturers

Thus far, our discussion of supply risk management has ignored strategic interactions among buyers. This omission is innocuous for mature industries, comprised of many small firms, where each one is a price-taker. However, there are many industries with only a few dominant firms. The interactions among such firms (e.g., Boeing and Airbus, Pfizer and Eli Lilly) are better described by oligopoly models, where actions of one firm affect other firms in the industry. In this section we will address two questions: (1) How does oligopolistic competition affect the supply risk management actions of buyers? (2) How do supply risk and supply risk management affect the oligopolistic competition among buyers?

The primary sources of insights for this section are Deo and Corbett (2008), Tang and Kouvelis (2009), and Wadecki et al. (2010). In all of these papers, the planning horizon has two stages. In the first stage buyers either (i) decide whether to enter the market (Deo and Corbett 2008), or (ii) select the structure of the supply network (Tang and Kouvelis 2009), or (iii) provide suppliers with financial subsidies (Wadecki et al. 2010). In the second stage, buyers compete. For the second stage, all three papers rely on the Cournot (1838) model[4] to capture competitive interactions among firms in the final product market. Deo and Corbett (2008) and Tang and Kouvelis (2009) model supply uncertainty as proportional random yield, which affects the output in the second stage of the model (i.e., when a manufacturer decides to produce $\bar{q}$ units, it successfully produces only $q = y\bar{q}$). In Wadecki et al. (2010) supply uncertainty is modeled via Bernoulli random variables and supply uncertainty is resolved in the first stage, before buyers engage each other in the final product market.

We begin with the discussion of Cournot competition game of the second stage.

[4] See Moorthy (1985) for applications of game theory to model competition: price and quantity competition, entry games, and dynamic oligopoly games.

14.5.1 SUPPLY RISK AND COURNOT COMPETITION

To help us appreciate the effect of supply risk on competition among firms, it is useful to recall insights from the classical Cournot (1838) model, to be used as a benchmark for subsequent results. There are N firms who compete by selecting quantities q_i they send to the market. If the total quantity in the market is $Q = \sum q_i$, the market price for the product is $p = a - bQ$. Firms incur unit production cost c. The Cournot equilibrium quantities, total output to the market, prices, and profits are $q^e = \frac{a-c}{(N+1)b}$, $Q^e = \frac{(a-c)N}{(N+1)b}$, $p^e = \frac{a+cN}{N+1}$, and $\pi^e = \frac{(a-c)^2}{(N+1)^2 b}$, respectively. To capture the consequences of competition in this classical model, we compare these equilibrium quantities with the optimal quantities in the centralized system. If all N firms were controlled by one decision maker, then effectively there is only one firm, $N = 1$. Its optimal monopolist production quantity (equal to total output to the market), resulting price, and optimal profit are $q^m = Q^m = \frac{a-c}{2b}$, $p^m = \frac{a+c}{2}$, and $\pi^m = \frac{(a-c)^2}{4b}$, respectively. Observe that the competitive system oversupplies, $Q^e = \frac{(a-c)N}{(N+1)b} > \frac{a-c}{2b} = Q^m$, which lowers the resulting market prices $p^e = \frac{a+cN}{N+1} < \frac{a+c}{2} = p^m$.

Next let us introduce supply risk in the form of random yield to the Cournot model, by assuming that the quantity that buyer i sends to the market is $q_i = y_i \bar{q}_i$, where $\bar{q}_i$ is the quantity that buyer i orders from its risky supplier. Assume that yields, $\{y_i\}_{i=1}^N$, are independent, identically distributed random variables, with mean μ, standard deviation σ, and coefficient of variation $\delta = \frac{\sigma}{\mu}$. Let d be the cost per unit of delivered product. In addition, assume that o is cost per unit of product ordered. One can compute the equilibrium order quantities, expected total market output, expected prices, and expected profits to be $\bar{q}^e = \frac{a-(o/\mu+d)}{\mu b(N+1+2\delta^2)}$, $EQ^e = \frac{[a-(o/\mu+d)]N}{b(N+1+2\delta^2)}$, $Ep^e = a - bEQ^e$, and $E\pi^e = \frac{[a-(o/\mu+d)]^2(1+\delta^2)}{b(N+1+2\delta^2)^2}$. To allow "apples to apples" comparisons with the classical Cournot model, we require that the cost per unit of product sent to the market in the model with supply risk, $o/\mu + d$, equals the equivalent cost c in the Cournot model. With this restriction, we observe that when yields $y_i \equiv 1$ (no supply risk, i.e., $\delta = 0$, $\mu = 1$), then the equilibrium quantities in the model with supply risk are equal to the equilibrium quantities in the classical Cournot model.

Comparing equilibrium quantities of the Cournot model and the Cournot model with supply risk, Deo and Corbett (2008) document an important effect of supply risk: with supply risk both the quantity ordered by the buyers and the total expected quantity sent to the market decrease. This reduction has significant societal ramifications: Deo and Corbett (2008) argue that it might be the culprit of chronic flu vaccine shortages in the U.S. If the coefficient of variation, δ, is high enough, then the order quantity under supply risk could be even smaller than the optimal order quantity of the monopolist buyer without supply risk.

We will continue the discussion of results in Deo and Corbett (2008) shortly, but before doing so, it is important to note an important insight from Tang and Kouvelis (2009) on the role of competition for a fixed number of buyers N. Tang and Kouvelis (2009) consider a Cournot game between $N = 2$ buyers, similar to Deo and Corbett (2008), but allow suppliers' random yields to be correlated. As the correlation between random yields increases, order quantities decrease and prices increase, but buyers' profits decrease. Intuitively, similar to what we have learnt in Section 14.4 on competing suppliers, buyer' products are differentiated by the states of nature when those products are available (i.e., states of nature where suppliers deliver). Greater correlation implies less differentiation and fiercer competition between buyers. Buyers that are eager to avoid competition should try to order from suppliers whose yields are uncorrelated. In particular, Tang and Kouvelis (2009) conclude that buyers, if possible, should not order from the same supplier (for this would introduce perfect correlation).

In this subsection we have discussed the effects of supply risk on the Cournot sub-game in the second stage of models in Deo and Corbett (2008) and Tang and Kouvelis (2009). Next we will study the games that happen in the first stage: market-entry competition, choice of supply chain structure, and financial subsidies games.

14.5.2 MARKET-ENTRY COMPETITION AND SUPPLY RISK

This subsection discusses insights from the market-entry competition game that determines the number of buyers N who will compete in the second-stage Cournot game of Deo and Corbett (2008). To ensure that N is finite, Deo and Corbett (2008) assume that each firm entering the market in the first stage pays a fixed entry fee f. They find that the effect of yield uncertainty on the number of entrants in equilibrium depends on how attractive the market is for the entrants. If the market size is relatively small, there will be fewer entrants in the model with supply risk compared to the model without supply risk. Interestingly, if the market size is large, limited supply risk could result in greater numbers of entrants in equilibrium. But even if the total number of entrants increases, the production of each of them decreases so much that the expected output in the model with supply risk is lower than the output without supply risk. Thus, even when the number of firms in the flu vaccine industry is determined endogenously, the effect of supply risk is to reduce the availability of the vaccine.

Deo and Corbett (2008) also consider a model where entry decisions are controlled by a central planner, but production quantities are determined competitively by the firms. The authors find that if the supply uncertainty is above a certain threshold, then the equilibrium number of firms in the market-entry game will be less than what the central planner would have chosen. Thus, a deregulated market, where firms themselves decide whether or not to start producing vaccine, could lead to greater vaccine supply risk, because with fewer firms there is less diversification.

14.5.3 SUPPLY RISK AND SUPPLY CHAIN STRUCTURE

Tang and Kouvelis (2009) assume that the number of buyers is fixed, $N = 2$, but in the first stage of the model, the buyers decide whether to sole source or dual source and with which supplier(s). First, the authors consider a model where

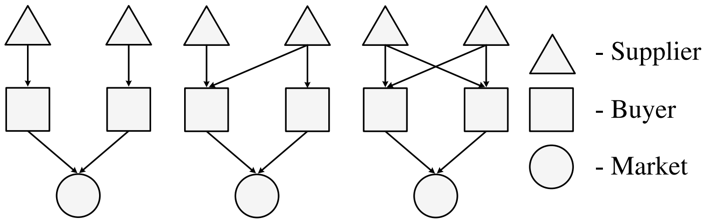


FIGURE 14.2 **Supply chain structures with two suppliers in Tang and Kouvelis (2009).**

only two suppliers are available, as illustrated in Figure 14.2, and characterize the equilibrium choice of sourcing strategy.

In this setting, the first trade-off a buyer faces is between the fixed cost of working with an extra supplier and benefits of diversification. Higher fixed cost retards diversification. Lower correlation between suppliers' random yields encourages diversification. This first trade-off is present in every model with a dual-sourcing decision. The second trade-off has features unique to problems with buyer competition: A buyer is choosing between costs of buyer competition and benefits of diversification. To reduce costs of competition, the buyers try to differentiate from each other. In particular, they would like to order from different suppliers. With only two suppliers available, the buyers end up choosing between sole sourcing from different suppliers (thus losing diversification benefits, but reducing costs of competition) and dual-sourcing from the same two suppliers (thus gaining diversification benefits, but increasing costs of competition between buyers).

Tang and Kouvelis (2009) also consider a model with two sets of two suppliers. One could interpret these sets as geographically separated supply bases. Therefore, the correlation of yields between sets is zero, while the correlation of yields within sets could be positive (or negative). The insights derived from the two trade-offs described above continue to hold.

14.5.4 SUBSIDIES TO SUPPLIERS IN THE PRESENCE OF BUYER COMPETITION

Wadecki et al. (2010) and Babich (2010) consider decisions of buyers to subsidize suppliers in the first stage of a planning horizon. These papers are motivated by the idea that ordering and payment policies of a buyer affect its supplier's financial viability, which in turn affects the supplier's ability to offer capacity to the buyer. These papers are inspired by numerous examples from the U.S. automotive industry, which has witnessed bankruptcies of not only automotive suppliers but also large manufacturers (both Chrysler and General Motors declared bankruptcy in 2009).

As Babich (2010) discusses, suppliers in or close to bankruptcy could face labor strikes, lose key personnel, reduce quality control efforts, or forego capacity investments. Creditors of bankrupt suppliers may try to "hold up" buyers, taking

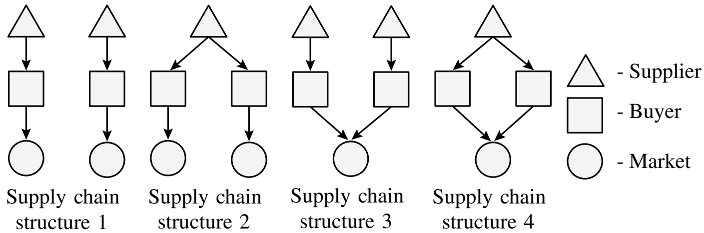


FIGURE 14.3 Four supply chain structures considered in Wadecki et al. (2010).

advantage of the critical role many suppliers play for buyers. A well-known example of a "hold-up" is UPF Thompson vs. Land Rover, discussed Section 14.7. A very recent example of a supplier's bankruptcy affecting its operational performance is the case of Chrysler and Plastech. After Plastech filed for bankruptcy in 2008, it stopped shipping plastic moldings to Chrysler and Chrysler had to shut down four plants.

The use of financial subsidies, without buyer competition, was studied by Babich (2010), who solves a dynamic procurement/subsidy problem and proposes conditions that allow procurement decisions to be made independently from subsidy decisions. Wadecki et al. (2010) solve the one-period variant of the problem in Babich (2010), but with buyer competition. They consider a model with $N = 2$ buyers and compare four supply chain structures (see Figure 14.3): (1) buyers do not compete and each has its own dedicated supplier, (2) buyers do not compete, but they share the same supplier, (3) buyers compete and each has its own dedicated supplier, and (4) buyers compete and share a supplier.

With supply chain structure 1, when deciding how much to pay its supplier, each buyer faces a key trade-off between increasing the probability that the supplier will be available to provide products in the second stage and the costs of raising money for a subsidy (as in Babich 2010).

The immediate effect of competition (supply chain structure 3) is that (unless the market size increases), subsidies to the suppliers decrease, because buyers now earn smaller profits by having to share the market and, therefore, are less keen to ensure that the suppliers stay in business.

According to Tang and Kouvelis (2009), buyers should always prefer supply chain structure 3 to supply chain structure 4, because by ordering from different suppliers, buyers can achieve differentiation and reduce competition. Wadecki et al. (2010) suggest that sometimes supply chain structure 4 might be preferable, due to a *cross-subsidy* effect—the shared supplier becomes very reliable because it receives subsidies from two buyers. Under supply chain structure 4, although each buyer pays less to the shared supplier than it would to a dedicated supplier, the total payment the supplier receives is greater in equilibrium. For example, because of the cross-subsidy effect, supply chain structure 2 always dominates supply chain structure 1 for the buyers.

Wadecki et al. (2010) offer another interpretation of the competition effect. Supply chain structure 3 might be preferable to supply chain structure 4 due to the *windfall profits* effect: When working with dedicated suppliers, buyers are betting on the other supplier going out of business and leaving the buyer to enjoy monopolist windfall profits. When the market size is large, the expected benefits of the windfall profits are greater than the cross-subsidy benefits. Thus, buyers prefer a fragmented supplier industry.

14.5.5 SECTION SUMMARY

The main insights from this section are as follows. Supply risk results in lower Cournot order quantities. The number of buyers who decide to enter the market could increase or decrease under supply risk, but the net effect of supply risk is that the expected quantity of products sold in the market always decreases, sometimes very drastically. Thus, one explanation for flu vaccine shortages in the U.S. could be the intrinsic uncertainty in the vaccine production process. Competition among buyers alters their risk management strategies. Without competition, when considering diversification buyers weigh fixed costs of working with an extra supplier against diversification benefits. With competition, buyers try to differentiate from each other, and sometimes they would rather forego diversification than order from the same set of suppliers. Ordering from the same supplier is not always bad, however: A supplier who receives multiple orders is more viable financially and, therefore, a more reliable supply source. Whether buyers prefer to have a few large shared suppliers or many small dedicated suppliers depends on the trade-off between cross-subsidy and competition effects.

14.6 Asymmetric Information

As discussed in the Section 14.1, in many settings suppliers are privileged with private information, for example, about the cost or reliability of their production or quality of items produced. Private information can affect the decisions and profits of the supply chain. In this section we discuss how a buyer facing supply disruption risks can (1) understand how asymmetric information affects the incentives of self-interested suppliers, (2) apply countermeasures against these incentives, and (3) understand when such countermeasures can change how the supply chain deploys traditional supply-risk management tools (e.g., backup production).

14.6.1 SUPPLIER MISREPRESENTATION INCENTIVES

Decisions are governed by incentives, and thus we begin this section by studying how suppliers' incentives are affected by their private information.

14.6.1.1 Scenario 1 - Private information about Production Cost. Consider a buyer who sole sources from a supplier with private information about its true production cost (which could be a function of its capacity utilization, order backlog, inventory level, etc.). How would the buyer figure out how much to offer to pay the supplier for production? Clearly the buyer could not ask the supplier to produce at a loss, but what is the supplier's break-even cost?

One can easily imagine that the supplier in Scenario 1 has an incentive to claim that its production is very costly, in an attempt to inflate the payment from the buyer. In other words, lower cost is of course better, but suppliers actually have an incentive to pretend to be worse (higher cost) than they actually are. The incentives in Scenario 1 are obvious, but incentives can also manifest themselves in less-obvious ways. Consider the following scenario.

14.6.1.2 Scenario 2 - Private Information about Supply Reliability. Suppose that a supplier has private information about its reliability, say the probability that it can successfully deliver the items ordered by the buyer. For example, unsuccessful delivery could correspond to items that are defective or destroyed in a factory fire. Casual intuition might suggest that a supplier would always have an incentive to appear more reliable, but as with Scenario 1 the opposite is true.

Suppose the supplier has to incur up-front expenses when working on the buyer's order and the buyer pays the supplier only when the supplier delivers. For the same contract from the buyer, a more reliable supplier earns higher expected profit. Therefore, to ensure that a supplier breaks even, the buyer has to pay more to a less reliable supplier. This creates an incentive for suppliers to pretend to be less reliable than they are.

In summary, self-interested suppliers can utilize their private information in an attempt to secure more favorable contracting terms from the buyer. In the next subsection we discuss countermeasures the buyer can use to manage this.

14.6.2 MANAGING SUPPLIER MISREPRESENTATION INCENTIVES

In this subsection we discuss a spectrum of approaches, applicable in the context of supply disruptions, to control supplier incentives to misrepresent their private information.

14.6.2.1 Take It or Leave It Offers. When the buyer faces a single supplier, the optimal negotiation format chosen by the buyer often involves making a take-it-or-leave-it offer to the supplier. By definition, when the buyer makes a take-it-or-leave-it offer it gives the supplier an ultimatum: accept my terms, or I walk away. This heavy-handed approach gives the supplier no opportunity to counter-offer and seek more favorable terms, thereby limiting the supplier's ability to leverage its private information. When contracting terms involve multiple dimensions (e.g. payment and non-delivery penalty), the buyer often finds it desirable to simultaneously offer the supplier multiple take-it-or-leave-it offers (e.g., high payment/high penalty or low payment/low penalty), and let the supplier select which offer it most prefers from the offer "menu."

14.6.2.2 Supplier Competition. Making a take-it-or-leave-it offer requires the buyer to credibly commit to not renegotiate with the supplier should the supplier choose to reject the buyer's offer (or menu of offers). The supplier, however, might not believe the buyer would really walk away from the deal. Without a credible walk-away threat, the buyer's take-it-or-leave-it offer is meaningless. When the buyer does not have the power or credibility to make such walk-away threats, supplier competition is an alternative countermeasure against supplier misrepresentation incentives. Regardless of the buyer's lack of supply chain power, a bid from one supplier can be used as a "stick" to gain concessions from other suppliers. Auctions have been used for millennia to buy and sell items and are a classic way to harness competition in a negotiation. The common process of putting a supply contract up for bid constitutes a "reverse" auction whereby bids are solicited from suppliers and the contract winner(s) and payment(s) are determined. When bidding, suppliers must trade off their desire for a large profit (e.g., asking a high price) against the chance that by being too greedy they might win less—or possibly none—of the buyer's business.

14.6.2.3 Power to design auction mechanisms. When the buyer has multiple suppliers and holds the power to make take-it-or-leave-it offers, it can leverage supplier competition and the ability to design a menu of offers. For instance, the buyer could employ a reverse auction in conjunction with a reserve price. The reserve price imposes an upper bound on the amount the buyer is willing to pay for the contract and thereby caps the profit of the winning supplier. More generally, supplier profits can be reduced further when the buyer has the power to dictate any kind of take-it-or-leave-it negotiation rules, for example, rules that favor weaker suppliers over stronger suppliers in order to enhance competition. As an example of the latter, if the buyer suspects that a particular supplier will enjoy a sizeable cost advantage over its opponents, she can impose an auction rule stating that, to win the business, this particular supplier must bid at least x dollars *below* the next-lowest bid. By "leveling the playing field" in this fashion, the buyer makes the stronger supplier bid lower just to stay competitive against weaker suppliers.

14.6.2.4 Other trading mechanisms. In general, the buyer's desire to manage supplier misrepresentation incentives can lead it to strategically employ take-it-or-leave-it offers, competition, or specific auction mechanisms. However, there may be reasons that the buyer cannot influence the choice of the trading mechanism, for example, due to historical precedence. It is therefore worthwhile to note the possibility of other trading mechanisms—for instance, a buyer and a supplier might make alternating counter offers ("bargain") until reaching an agreement, and thereby split the supply chain profits in some exogenously defined way.

In the extant literature on decentralized supply risk management in the presence of asymmetric information, all of these various approaches are studied. Lim (1997), Yang et al. (2009a), Yang et al. (2009b), Chaturvedi and Martínez-de-Albéniz (2010), and Wan and Beil (2008) all assume a strong buyer able to make take-it-or-leave-it offers. Wan and Beil (2008) examine a spectrum of buyers, ranging from those able to make take-it-or-leave-it offers to those who must rely solely on competition for price concessions. Gurnani and Shi (2006) examine a bargaining setting.

14.6.3 EFFECT OF ASYMMETRIC INFORMATION ON SUPPLY RISK MANAGEMENT TOOLS

Countermeasures to mitigate supplier incentives are not deployed in a vacuum. For better or worse, they often affect how the supply chain operates. For example, when the buyer makes take-it-or-leave-it offers to suppliers, there is a chance that suppliers will reject the offer, leaving the buyer empty-handed. In the context of supply risk, an important consideration is how such countermeasures affect the buyer's use of supply risk management tools, such as: backup production at the buyer, backup production at the supplier, nonperformance penalties, quality screening, supplier qualification screening, dual sourcing, supply base diversification, etc. We explore these issues in this subsection.

14.6.3.1 Supplier/Buyer Backup Production. As discussed in the Section 14.1, it is plausible that the supplier or buyer could utilize backup production in the event of a supply disruption. Yang et al. (2009a) study such a setting in which the buyer faces a single supplier possessing private information about its reliability. It is found that when the supplier's reliability is its private information, the buyer is less likely to use the backup production option of the supplier, but more likely to rely on its own (more costly) backup option. Why does asymmetric information about supply risk cause the buyer to utilize supplier backup production less? If the supplier is asked to use its backup production in the event of a disruption, the cost differential between a more reliable supplier and a less reliable supplier grows further, because the latter is more likely to suffer a disruption and, hence, more likely to incur the cost of using backup production. Consonant with the discussion in section 14.6.1 for Scenario 2, this widening of the cost gap increases the more reliable supplier's incentive to misrepresent itself. Thus, the buyer may choose to forego the supplier backup production option to reduce the high-reliability supplier's misrepresentation incentive.

14.6.3.2 Supplier Diversification. Supplier diversification is another tool that the buyer can use to mitigate supply risks. Yang et al. (2009b) study a buyer facing two suppliers that are each subject to random supply disruption. To increase the chance of delivery the buyer can diversify, that is, contract with both suppliers simultaneously. Each supplier's reliability (probability of disruption) is their own private information. The buyer sets payment-on-delivery terms to ensure that the suppliers have an incentive to deliver. Following the theme of this section, such contingent payments create an incentive for suppliers to misrepresent their reliability. Because asymmetric information effectively makes diversification more costly and competition more attractive, the buyer utilizes diversification less. Chaturvedi and Martínez-de-Albéniz (2010) show a similar shift away from diversification (caused by asymmetric information about production costs).

14.6.3.3 Supplier Qualification Screening. As described in the Introduction, buyers often rely on supplier qualification screening to reduce supplier nonperformance risks. Since suppliers have private information about their costs, buyers often use auctions to mitigate suppliers' incentives to inflate their prices. In practice, qualification screening typically precedes the auction, meaning that the buyer wastes time and money qualifying suppliers who wind up losing the auction. Wan and Beil (2008) find that this practice can be improved upon by postponing all or part of the supplier qualification screening until after the auction. In particular, when qualification is costly, reducing qualification costs through judicious post-qualification can more than offset expected increases in the contracting costs resulting from higher prices in the auction (which arise since postponing qualification screening means that some attractive auction bids might be disqualified).

14.6.3.4 Efforts on Quality. Defective parts are costly for the supply chain, requiring either rework if they are discovered by the buyer, or—even worse—warranty and goodwill costs if they are discovered by the end user (the buyer's customer). Lim (1997) focuses on a buyer unsure of the quality level of her supplier, namely the probability that any given part from the supplier will be defective. While the buyer would like to share quality-related costs with the supplier, the asymmetric information makes it difficult for the buyer to decide how much to penalize the supplier for defects caught at inspection and in the field. Intuitively, high penalties make the contract very unattractive to an unreliable supplier. Lim (1997) finds that to ensure the contract is attractive to any supplier type—even the least reliable suppliers—the buyer winds up absorbing more of the quality-related costs.

In a similar vein, Baiman et al. (2000) study a situation where the supplier's effort to reduce the fraction of defective units in production or the buyer's effort to inspect incoming units may not be publicly observed and thus may not be contractible. The authors find that the optimal channel profit can still be obtained if the buyer and supplier can contract upon one of the following three sets of information: (a) the supplier's defect prevention effort; (b) the buyer's quality inspection effort and product failures reported by either the buyer or the customer; and (c) product failures reported by both the buyer and the customer.

In Babich and Tang (2010), it is asymmetric information about the supplier's potential action to adulterate the product that creates supply risk for the buyer in the first place. As numerous recent examples of product recalls show, when buyers cannot monitor their overseas suppliers, those suppliers, under tremendous pressure to cut cost, are tempted to compromise on quality. Babich and Tang (2010) discuss various approaches to mitigating adulteration risk. Unfortunately, many traditional tools can be ineffective in practice. Supplier certification, (e.g., ISO9000), only guarantees that the supplier has the right process in place, but it does not guarantee that the supplier follows it. Product inspection at the time of delivery from the supplier may work, but, some suppliers deliberately try to fool inspection agencies (e.g., the reason melamine is a culprit in recent milk and pet food recalls is that melamine increases protein count on the tests of products and thus those products appear normal). Supplier liability and warranties are difficult to enforce with overseas suppliers, who are subject to different legal systems, and frequently do not have resources to compensate the buyer for the losses. An approach that seems promising and that is studied in Babich and Tang (2010) is the deferred payment mechanism, when the buyer defers the payment to the supplier for a prespecified period, and pays contingent on the customers not discovering defects. An example of a deferred payment mechanism is a popular financial contract, called trade credit. Babich and Tang (2010) also study an inspection mechanism, when the payment to the supplier is contingent on the product passing inspection, and the combined mechanism, which is a combination of the inspection and the deferred payment mechanisms. Surprisingly, Babich and Tang (2010) prove that the combined mechanism is redundant. They identify four factors that determine the dominance of deferred payment mechanism over the inspection: (a) the inspection cost relative to inspection accuracy, (b) the buyer's liability for adulterated products, (c) the difference in financing rates for the buyer and the supplier relative to the defect-discovery rate by customers, and (d) the difference in production costs for adulterated and unadulterated product. Interestingly, the deferred payment mechanism is preferred by the buyer if the gap in costs of producing unadulterated and adulterated products is low and if the buyer's liability for adulterated products is low.

14.6.3.5 Nondelivery penalties. Yang et al. (2009a) examine nondelivery penalties and whether or not the buyer wishes to set the penalty high enough to induce the supplier to use its backup production option in the event of a supply disruption. They find that the buyer may choose lower penalties under asymmetric information because, under asymmetric information, the buyer's incentive payment to the supplier is tied to the penalty value. Hence, the buyer would always choose the lowest penalty value that induces the desired action by the supplier. An interesting implication of this observation pertains to the buyer's ability to transfer supply-shortage risk to its supplier. Under symmetric information, the buyer can optimally set the penalty to be equal to the value of lost revenue in case of supply disruption and thus completely isolate itself from supply risk. Under asymmetric information, this is no longer optimal. Therefore, under asymmetric information the buyer may still have to carry supply disruption risk.

Nondelivery penalties also feature prominently in Gurnani and Shi (2006), where a buyer and a supplier have differing estimates of the supplier's reliability. While the buyer knows the supplier's reliability self-estimate and the supplier knows the buyer's estimate of the supplier's reliability, neither the buyer nor supplier knows the true reliability for sure. Unlike the papers described above where the buyer essentially seeks to discover information held privately by the supplier, in Gurnani and Shi (2006) the buyer's beliefs about reliability are not affected by knowing the supplier's self-estimate. Depending on whether the buyer's estimate is larger or smaller than the supplier's, the paper suggests employing contract terms incorporating either downpayment or nondelivery penalty.

14.6.3.6 Value of Supply Risk Management Tools under Asymmetric information. Thus far we have seen that asymmetric information effectively makes it more expensive to use supplier backup production, diversification, and supplier qualification screening. However, this does not mean that such tools are no longer valuable, as we discuss next.

Yang et al. (2009a) find that asymmetric information about supplier reliability causes supplier backup production to be used less, but they also show that the value of supplier backup production for the buyer is not necessarily larger when it perfectly knows the supplier's reliability. In particular, adding a cheap backup production option diminishes the supplier's benefit of misrepresenting its reliability (since reliability overall becomes less of a concern). This incentive reduction provides an extra benefit which does not exist when information is symmetric.

Although Yang et al. (2009b) find that asymmetric information about supplier reliability causes the buyer to utilize diversification less, they also show that having a dual-sourcing option (i.e., having two potential suppliers) is very valuable for the manufacturer even if she does not use this option to diversify her supply. Merely having two suppliers allows the buyer to play one supplier against the other to receive better pricing, a competition benefit which is absent when the buyer has perfect information about the suppliers.

In Wan and Beil (2008), supplier qualification screening is still imperative—a buyer would be loathe to enter into an important contract with a supplier until knowing the supplier has passed qualification screening. While qualification screening is still extremely valuable, the manner in which it is used changes due to asymmetric information. In particular, post-qualification can be part of an optimally balanced supplier qualification strategy, which may become ever more important as supply sources globalize and asymmetry of information worsens.

14.6.4 SECTION SUMMARY

Self-interested suppliers can misrepresent their private information to the buyer, for example, by misreporting their reliability. The desire for more favorable contract terms (e.g., a higher price or lower nonperformance penalties) can provide suppliers an incentive to do so. To mitigate this incentive while managing supply risks in a decentralized supply chain, the buyer can deploy optimally designed take-it-or-leave-it-offers, supplier competition, or both. These countermeasures affect—and are affected by—the supply risk management tools available to the buyer. For example, the tools themselves can change how the buyer deploys countermeasures. For instance, a buyer with very cheap backup production (at the buyer or supplier) worries little about reliability and hence need not worry greatly about imposing stringent countermeasures against supplier incentives, and a buyer with a dual-sourcing option might choose to use it to reduce incentives via competition (taking cost-reduction benefits) rather than diversification (taking supply-risk-reduction benefits).

14.7 Conclusions

Research on decentralized supply risk management is relatively new, yet it has already produced a number of interesting, surprising, and salient insights. We believe that more exciting findings lie ahead in this field, and conclude this chapter by discussing several promising future research directions.

In traditional risk management, statistical data analysis is a crucial step when forming practical recommendations. Yet the papers discussed in this chapter are based primarily on theoretical analysis. We think that an important future research direction is to adopt a more data-driven approach. Such data-driven research can be in the form of empirical research, which would take advantage of increasing amounts of data about the causes, durations and costs of supply disruptions (e.g., Hendricks and Singhal 2003) to validate theoretical recommendations and to quantify the value of various tools for decentralized risk management. Alternatively, data-driven research can take the form of experimental studies, which could explore the effects of managerial attitudes toward supply risk (e.g., Schweitzer and Cachon 2000) and the interactions between multiple players in decentralized systems in experimental settings (see Trading Agent Competition, `http://www.sics.se/tac`).

As was highlighted in this chapter, a key form of decentralized supply risk management is information asymmetry, for example, suppliers having private information about their costs and reliabilities, or buyers having better information about the market value of their products. A promising direction for future research is to utilize richer models of beliefs held by parties regarding others' private information. Buyers' beliefs about suppliers' reliabilities may differ from suppliers' beliefs about each other. Such differences in beliefs raise questions about who in the supply chain has higher-quality information, and what is the best way to delegate supply-chain decisions (e.g., evaluating the benefit of using a Procurement Service Provider, see Accenture 2003). In addition, it would be helpful to analyze the robustness of supply chain decisions to mistaken beliefs. Alternatively, rather than enriching Bayesian models of beliefs, one can adopt an entirely different approach and explore non-Bayesian, belief-free models of information asymmetry as in Sharma et al. (2009). The presence of asymmetric information also can cause moral hazard problems in decentralized supply risk management. For example, when a buyer commits to subsidizing a risky supplier, the buyer might unintentionally be encouraging the supplier to take further risks that are not in the buyer's best interest.

An interesting direction for future research is to analyze cooperative decision-making in the presence of supply risk. For example, suppliers may collude, thus changing the buyer's risk profile. Likewise, a group of buyers may cooperate to make group purchases, which would then lead to questions about how they should allocate the supply risk among themselves. Furthermore, suppliers and buyers may cooperate to make the supply chain more resilient to risks as long as both parties benefit from such concerted efforts.

For methodological approaches to supply risk management, an important distinguishing characteristic of the supply risk is whether or not it is priced by financial markets. A related practical question is whether financial instruments can be used to manage supply risks, and how such instruments would interact with traditional operational tools such as backup production and multisourcing. If financial replication is not possible, perhaps actuarial risk measurement and management can be used instead. Likewise, the relatively more mature field of quality control may provide inspiration for managing supply risks in decentralized networks (e.g., see Zipkin 2009).

Much of the extant research literature focuses on supply risk as it applies to quantity, (i.e., an order not being filled in full). Identifying and studying other forms of supply risk would be a valuable research direction. For example, one important type of supply risk, delayed delivery of an order, is understudied and worthy of further consideration. Such delays are particularly relevant in decentralized environments, where a buyer's ability to meet its own deadlines may require successful coordination of deliveries from several independent suppliers.

Another connection that is worthy of further research is the one between risk management and product design. A manufacturer's product design is often closely linked to its supply chain configuration, which in turn influences the manufacturer's supply risk. For example, one could investigate the risk implications of using modular products, which require the final assembler to put together a few modules built by subassemblers, versus nonmodular products, which require the final assembler to combine a large number of parts and components. The former design type leads to long supply chains with fewer suppliers at each echelon, whereas the latter design type leads to shorter supply chains with many suppliers at each echelon (Bernstein and DeCroix 2004, Wang et al. 2009). Such different supply chain structures have a significant effect on the final assembler's vulnerability to supply risks.

Finally, as discussed in Section 14.1, there are myriad causes of supply disruptions, and future research may move beyond generic models of supply risk to look deeper into specific causes of supply disruptions. For example, a plant fire and an impasse in labor negotiations may be similar as far as their end results are concerned—both will cause a temporary halt in production. However, the preparedness and recovery plans for a disruption caused by a plant fire may be significantly different from the plans that are in place for labor negotiations. High-fidelity models, which capture considerations that are relevant to specific causes of supply risks, may be used for prescriptive purposes in supply risk management.

REFERENCES

Accenture. 2003. Executive insights into the growing use of procurement outsourcing. 2003 procurement survey.

Babich V. 2006. Vulnerable options in supply chains: Effects of supplier competition. *Naval Research Logistics*, **53**(7):656–673.

Babich V. 2010. Independence of capacity ordering and financial subsidies to risky suppliers. *Manufacturing and Service Operations Management*, Published online before print February 16, 2010, DOI:10.1287/msom.1090.0284.

Babich V., Burnetas A.N., Ritchken P. H. 2007. Competition and diversification effects in supply chains with supplier default risk. *Manufacturing and Service Operations Management*, **9**(2):123–146.

Babich V., Tang C.S. 2010. Managing opportunistic supplier product adulteration: Deferred payments, inspection, and combined mechanisms. Working paper. Washington, DC: McDonough School of Business, Georgetown University.

Baiman S. Fischer P.E., Rajan M.V. 2000. Information, contracting, and quality costs. *Management Science*, **46**(6):776–789.

Bernstein, F., DeCroix G.A. 2004. Decentralized pricing and capacity decisions in a multitier system with modular assembly. *Management Science*, **50**:1293–1308.

Bertrand J. 1883. Theorie mathematique de la richesse sociale. *Journal des Savants*, **67**:499–508.

Biddle F.M. 1997a. Boeing effort to cushion itself from cycles backfires—production overhaul hinders plane maker's ability to handle order boom. *The Wall Street Journal (Eastern edition)*, **October 24,** 1.

Biddle F.M. 1997b. Boeing to miss 12 deliveries in September. *The Wall Street Journal (Eastern edition)* **September 16,** A.2.

Casey N. 2007. Mattel issues third major recall; top toy brands Barbie, Fisher-Price are latest facing lead-paint issues. *The Wall Street Journal (Eastern edition)* **September 5,** A.3.

Chaturvedi A., Martínez-de-Albéniz V. 2010. Optimal procurement design in the presence of supply risk. *Manufacturing and Service Operatiopn Management*, **13**(2):227–243.

Ciarrallo F.W., Akella R., Morton T.E. 1994. A periodic review, production planning model with uncertain capacity and uncertain demand—optimality of extended myopic policies. *Mannagement Science*, **40**(3):320–332.

Costello M. 2008. Shipping insurance cost soars with piracy surge off Somalia. *The Times*, **September 11**.

Cournot A. 1838. *Recherches sur les Principes Mathematiques de la Theorie des Richesses*. Paris, France: Hachette.

Deo S., Corbett C. 2008. Cournot competition under yield uncertainty: The case of the US influenza vaccine market. Working paper, Evanston, Illinois: Kellogg School of Management, Northwestern University, and Los Angles, California: UCLA Anderson School of Management, UCLA.

Federgruen A., Yang. N. 2009. Competition under generalized attraction models: Applications to quality competition under yield uncertainty. Working paper. New York, New York: Graduate School of Business, Columbia University.

Froot K.A., Scharfstein D.S., Stein. J.C. 1993. Risk management: Coordinating corporate investment and financing policies. *Journal of Finance*, 1629–1658.

Gurnani H., Shi M. 2006. A bargaining model for a first-time interaction under asymmetric beliefs of supply reliability. *Management Science*, **52**(6):865–880.

Hendricks K.B., Singhal V. R. 2003. The effect of supply chain glitches on shareholder wealth. *Journal of Management*, **21**:501–522.

Jennings D. 2002. The best law X 4 X far? *Supply Management*, **7**(6):40–41.

Karnani A. 1985. Strategic implications of market share attraction models. *Management Science*, **31**(5):536–547.

Lim, W.S. 1997. Producer-supplier contracts with incomplete information. *Management Science*, **47**(5):709–715.

McCracken J. 2006. Ford gets cut off by top supplier in push against detroit squeezes. *The Wall Street Journal (Eastern edition)* **October 18**, A3.

Moorthy K.S. 1985. Using game theory to model competition. *Journal of Marketing Research*, **22**(3):262–282.

Morse A. 2006. Laptop makers may seek redress from sony due to battery recall. *The Wall Street Journal (Eastern edition)*, **October 17,** B.3.

Myers R. 2007. Food fights. *CFO*, **June**, 70–74.

Nishiguchi T., Beaudet A. 1998. The Toyota group and the Aisin fire. *Sloan Management Review*, **40**(1):49–59.

Papadakis I.S. 2003. On the sensitivity of configure-to-order supply chains for personal computers after component market disruptions. *International Journal of Physical Distribution and Logistics Management*, **33**(10):934–950.

Savage M. 1999. Component aftershock hits channel. *Computer Reseller News* **October 18, 1999**(864) 166.

Schweitzer, M.E., Cachon G.P. 2000. Decision bias in the newsvendor problem with a known demand distribution: Experimental evidence. *Management Science*, **46**:404–420.

Sharma S., Babich V., Teneketzi D., Van Oyen M. 2009. A decentralized mechanism implementing in Nash equilibria the optimal centralized solution of a supply-chain problem. Working paper, University of Michigan.

Sheffi Y. 2005. *The Resilient Enterprise: Overcoming Vulnerability for Competitive Advantage*. Cambridge, Massachusetts:The MIT Press.

Swinney R., Netessine S. 2009. Long-term contracts under the threat of supplier default. Manufacturing and Service Operations Management, **11**(1):109–127.

Tang S.Y., Kouvelis P. 2009. Supplier diversification strategies in the presence of yield uncertainty and buyer competition. Working paper. Coral Gables, Florida:University of Miami.

Tomlin B., Wang Y. 2010. Operational strategies for managing supply chain disruption risk. In: Kouvelis P., Boyabatli O., Dong L., Li R., editors, *Handbook of Integrated Risk Management in Global Supply Chains*. New York: Wiley.

Wadecki A., Babich V., Wu O. 2010. Manufacturers' competition and subsidies to suppliers. Gurnani H., Mehrotra A., Ray S., editors, *Managing Supply Disruptions*. London: Springer-Verlag.

Wall Street Journal. 2007. Spinach recall weakens trust in vegetable safety. *The Wall Street Journal (Eastern edition)* **February 5,** B.2.

Wan Z., Beil D. R. 2009. RFQ auctions with supplier qualification screening. *Operations Research,* **57**:934–949.

Wan Z., Beil D.R. 2008. Bargaining power and supply base diversification. Working paper, Ann Arbor, Michigan: Ross School of Business, University of Michigan.

Wang H., Aydin G., Hu J. 2009. A complexity model for assembly supply chains in the presence of product variety and its relationship to cost. Working paper. Ann Arbor, Michigan: University of Michigan.

White G. L. 2000. Unruly element: Russian maneuvers are making palladium ever more precious—they force Detroit to pay soaring prices for a metal that's used in most cars—tapping a mine in Montana. *The Wall Street Journal (Eastern edition)* **March 6,** A.1.

White, Joseph. 2005. Ford to pay up to $1.8b on Visteon. *WSJ* **May 26** A3.

Wu Z., Choi T.Y. 2005. Supplier–supplier relationships in the buyer–supplier triad: Building theories from eight case studies. *Journal of Operations Management,* **24**(1):27–52.

Yang Z., Aydın G., Babich V., Beil R. D. 2009a. Supply disruptions, asymmetric information, and a backup production option. *Management Science,* **55**(2):192–209.

Yang Z., Aydın G., Babich V., Beil D.R. 2009b. Supply disruptions, asymmetric information, and a dual-sourcing option. Working Paper. Ann Arbor, Michigan: Department of Industrial and Operations Engineering, University of Michigan.

Yano C. A., Lee H.L. 1995. Lot sizing with random yields: a review. *Operations Research,* **43**(2):311–334.

Zimmer K. 2002. Supply chain coordination with uncertain just-in-time delivery. *International Journal of Production Economics,* **77**(1):1–15.

Zipkin P. 2009. Quality snags in the mortgage-finance supply chain Working paper. Durham, North Carolina: Duke University.

Chapter Fifteen

Using Supplier Portfolios to Manage Demand Risk

VICTOR MARTÍNEZ-DE-ALBÉNIZ

In planning production for products with uncertain demand, companies can use portfolios of supply contracts to mitigate demand risk. In such a portfolio, a buyer contracts supply from two different sources: one with a lower total cost, but no adjustment flexibility; and a more expensive one, with which the buyer can increase or decrease production after receiving better demand forecasts. Thus, through the portfolio, it is possible to serve volatile demands at a lower cost. This chapter presents two models where supplier portfolios can be optimized. The first model analyzes how to build a supply options portfolios in a static environment, at the beginning of a new selling season. The second model considers a dynamic setting where the buyer, before the beginning of the season, receives updated demand forecasts, and sequentially places orders with the suppliers that are fast enough to deliver before sales start. These models capture the basic trade-offs involved in multisourcing in many industries.

Handbook of Integrated Risk Management in Global Supply Chains, First Edition. Edited by Panos Kouvelis, Lingxiu Dong, Onur Boyabatli, and Rong Li.

15.1 Introduction

One of the challenges in supply chain planning is to identify and mitigate demand risk, through appropriate supply decisions that avoid over-production and shortages. Matching supply and demand is difficult since supply decisions usually require a long time to be implemented and realized demand often deviates from its forecast. This task is especially intricate when a firm plans the production of new products, for which demand uncertainty tends to be very high. In some industries, such as electronics or fashion apparel, this is the norm, as products have a very short life cycle. In such uncertain environments, ordering the right quantities to serve demand is critical for a firm's long-term survival.

Since the quality of demand forecasts improves with time, any opportunity to delay ordering decisions is valuable. On the other hand, order postponement is typically associated with higher costs: Production and delivery must be expedited, and there is less room for process optimization. Thus, the buyer can typically achieve either lower costs but with a higher demand risk, by working with a supplier that requires advance commitment; or it can reduce the chances of supply-demand mismatch, at a higher cost, by working with a supplier that allows last-minute ordering.

Interestingly, when fixed ordering costs are very small compared to variable (per-unit) ordering costs, the buyer does not need to choose between low-cost/low-flexibility or high-cost/high-flexibility sourcing. In fact, by working with both types of suppliers at the same time, the buyer obtains the best of both worlds. Specifically, first it can sign a contract with the supplier that offers the lowest total cost for a portion of demand that is very likely to materialize. Hence, since the buyer's exposure is limited, it is ready to commit long in advance for such a contract. In addition, it can delay the ordering of the remaining units until more accurate demand forecasts are obtained. The postponement of this decision can be implemented by either working with a short lead-time supplier, or by arranging a flexible contract that allows setting the final ordering quantity after demand is realized.

A few industries have adopted such purchasing practices, mainly in the last decade. In the electronics industry, where demand is quite volatile, Hewlett-Packard (HP) has developed a Procurement Risk Management (PRM) program to build supply portfolios, see Nagali et al. (2008). HP has applied PRM to direct components such as memory, hard disk drives, plastics or even custom integrated circuits, for a total expenditure of $7 billion in 2006. Through the PRM group, HP builds a portfolio of supply contracts from its suppliers. The portfolio usually contains a fixed quantity contract that just covers the demand in the most pessimistic scenario, as well as a flexible quantity contract that allows HP to decide on the appropriate supply volume after observing realized demand. PRM not only allows to control demand risks, but can also manage material cost risks.

Other examples of the use of supply portfolios can be found in apparel retailing. The German retailer Adidas in some occasions uses two suppliers for a particular product: one in East Asia and another one in Germany. A large order is placed with the Asian supplier. In addition, if the demand is higher than expected, and no additional shipments are planned, then Adidas places a rush order with the local supplier, which is more expensive but allows the retailer to avoid stocking out. Such a strategy provides better reactivity to demand variability (through local quick-response supply) at a low cost (since most of the volume is sourced from a low-cost country).

Also in apparel retailing, in Spain, Friday's Project specializes in designing and selling fashion products. Production is subcontracted to the Far East and Europe. When a new design is finalized, the usual procedure is to place a base order at an Asian supplier, for less than the expected demand. Later on, if the item sells well, the company places an additional order from the local supplier to ensure sufficient supply. This approach allows Friday's Project to significantly reduce demand risk in this very volatile industry.

Using supply portfolios requires significant changes, not only for the buyers, but also for the suppliers involved. Indeed, buyers need to become more sophisticated in their purchasing processes. They have to continuously update their demand forecasts to identify when it is necessary to place additional orders, or exercise existing contracts. They have to be able to effectively communicate their needs for flexibility to suppliers, and price them correctly. On the other hand, suppliers need to be prepared to quickly react to the buyer's requests, with agile production and delivery processes, such as delayed differentiation and expediting.

A number of new challenges have appeared with the use of supply portfolios.

- At an operational, day-to-day level, buyers have to decide how to coordinate deliveries from multiple suppliers. Effective inventory management policies are needed.
- At a tactical level, buyers must know how to install supply capacity at different suppliers. This requires evaluating the trade-offs between cost and the subsequent flexibility derived from the capacity. Understanding capacity costs, with possible set-up charges, economies of scale or quality control requirements, is critical.
- At a strategic level, buyers need to be aware of the repercussions of portfolio purchasing on suppliers and industry dynamics. Indeed, in the long term, suppliers may change their pricing policies to reflect the value they create for the buyer: A flexible supplier might be able to increase its price without losing volume (captive buyer), while inflexible suppliers might enter price competition that reduce long term cost for the buyer. Understanding such dynamics is key before a buyer decides to use portfolios.

The objective of this chapter is to provide simple guidelines on how to use portfolios, focusing on the operational and tactical level. For this purpose, we focus on the case of a buyer that obtains supply from two suppliers[1]. We provide two models to capture the advantages of a portfolio approach, compared to single sourcing.

First, we develop a static model that captures the main trade-offs in the PRM program at HP. Initially, the buyer has a demand forecast and based on this information it reserves production capacity at the suppliers. The suppliers offer a pricing structure that provides the buyer with some flexibility regarding final production orders (i.e., the final order might be anywhere between zero and the contracted capacity). Later, the buyer observes the realized demand and places final orders. We analyze the buyer's capacity decision, and show that the optimal portfolio capacity investment is found by solving two critical fractile equations. This model is based on a single-period version of Martínez-de-Albéniz and Simchi-Levi (2005).

Second, we develop a dynamic model that captures the ordering dynamics in apparel retailing. As a new design is completed, the buyer has an initial demand forecast, that will continuously improve as we get closer to the selling season. Two suppliers, with different prices and lead-times, are available. The buyer first places an order with the longer lead-time supplier, and then an order with the shorter lead-time supplier, after receiving a demand update. We characterize how to sequentially place the orders. The decision again involves solving critical fractile equations. This model is based again on Martínez-de-Albéniz and Simchi-Levi (2005), but uses the demand model of Calvo and Martínez-de-Albéniz (2010).

These two models make different assumptions on how the ordering process works, but are conceptually similar. In the static model, the portfolio is advantageous since each supplier has a different cost structure to provide flexibility. In the dynamic model, supplier flexibility is modeled through lead time. Using a portfolio allows to reduce the order from the long lead-time supplier and only request additional supply from the short lead-time supplier when the forecast update reveals that demand is high.

We next review the relevant academic literature on the topic. We then present the two models, and conclude with a discussion on some open research questions related to multiple sourcing.

15.2 Literature Review

The academic literature on multiple sourcing is quite extensive. An overview specializing on inventory management models is provided by Minner (2003). At least three different research streams exist, depending on how supplier differences are modeled: (i) suppliers differ in price and lead-time, (ii) suppliers differ in price and reliability, (iii) suppliers differ in price and flexibility. The static model in this chapter would fall in (iii), while the dynamic model would be in (i).

[1] The models presented have two suppliers to simplify exposition, but all the results extend to the case of $n > 2$ suppliers.

A number of papers analyze inventory management policies with multiple sources with different lead-times. In a seminal paper, Fukuda (1964) showed that when the lead-time difference between the suppliers is equal to 1, then a double base-stock policy is optimal for the buyer. However, when the lead-time difference is greater than 1, the optimal replenishment policies do not exhibit the base-stock property anymore, see Whittemore and Saunders (1977) and Feng et al. (2005). Double base-stock policies have nevertheless been used as heuristics, see for instance Moinzadeh and Nahmias (1988), Moinzadeh and Schmidt (1991) or Tagaras and Vlachos (2001). Veeraraghavan and Scheller-Wolf (2008) optimize the parameters of this policy in a capacitated setting. Martínez-de-Albéniz (2005) uses double base-stock policies to analyze price competition between two suppliers. All these models analyze multiperiod models, and this is why the structure of the optimal policy is difficult to obtain. In contrast, Song and Zipkin (2007) consider a single sales-period model, where inventory decisions can be adjusted over time, as the demand forecast is updated. Calvo and Martínez-de-Albéniz (2010) use a conceptually similar model where suppliers are used sequentially, and derive optimal sourcing decisions together with supplier equilibrium pricing. This approach hence simplifies the problem into a standard multiperiod inventory problem with one source.

Much work has also been done on multiple sourcing with unreliable suppliers. Some papers have modeled unreliability as uncertainty in production yield. In other words, suppliers deliver only a fraction of the orders that were placed, because the production line is not perfectly reliable. In that case, supplier diversification is beneficial in order to reduce yield uncertainty. This approach is motivated by quality problems or response-time uncertainties. For example, Gerchak and Parlar (1990), Anupindi and Akella (1993), and Parlar and Wang (1993) discuss the optimal diversification policy for the buyer (i.e., its optimal replenishment strategy, with two suppliers). Dada et al. (2007) and Federgruen and Yang (2008) analyze the general problem with multiple suppliers, in a newsvendor context. Some other papers have considered unreliability as uncertainty in production capacity. In the extreme case, when there is a chance that a supplier has zero capacity to offer, using dual sourcing reduces the risk of a supply disruption. Tomlin (2006) studies contingency strategies, one of them being multisourcing. Babich et al. (2007) analyze optimal sourcing and the effect on supplier competition. Chaturvedi and Martínez-de-Albéniz (2008) design optimal procurement mechanisms when suppliers might suffer disruptions. Chaturvedi and Martínez-de-Albéniz (2009) optimize capacity decisions when disruptions are possible, and particularly in a multisourcing setting.

Finally, some papers focus on sourcing diversification when suppliers offer flexibility through their cost structure. Most of the time, the cost structure consists of a per-unit capacity reservation fee, plus a per-unit execution fee for each unit delivered. In other words, suppliers offer quantity option contracts to the buyer. Sometimes, a spot market also exists that can provide supply without previous capacity reservation from the buyer. Wu et al. (2002) consider the reservation problem with one supplier. They assume that the demand is decided by the buyer, by maximizing its utility function, and determine the optimal capacity to be reserved. A multi-sourcing version of this approach is presented in Wu and Kleindorfer (2005). Suppliers are characterized by their execution unit cost and their available capacity, and offer option contracts to the buyer. Due to the modeling assumptions, suppliers can be sorted so that the buyer's optimal reservation strategy is to use some of the highest ranked, up to their available capacity. In contrast, Martínez-de-Albéniz and Simchi-Levi (2005) consider a stochastic demand, and develop a multiperiod model with multiple suppliers and a spot market, where portfolios of options can be optimized. Martínez-de-Albéniz and Simchi-Levi (2006) solve a similar problem with a risk-averse buyer. Fu et al. (2010) explore the benefits of using portfolios, and provide optimization algorithms. Martínez-de-Albéniz and Simchi-Levi (2009) analyze suppliers' competition when the buyer uses an options portfolio, and characterize equilibrium option prices.

15.3 A Static Model

We first introduce a simple model of multiple suppliers sourcing based on capacity options.

Consider one buyer that sells a product at a price p per unit, in a single sales period. The demand for the product is stochastic, denoted D. The probability density function (p.d.f.) of D is denoted f and its cumulative distribution function (c.d.f.) F. We let $\bar{F} = 1 - F$.

In order to fulfill the demand, the buyer must order the product from two suppliers, at a cost. We assume that any unfulfilled demand is lost, and as a result, the buyer needs to balance the ordering cost from the suppliers and the potential loss of revenue, by adjusting its sourcing strategy.

In this static model, the suppliers are assumed to differ in their price and flexibility. The price/flexibility level of each supplier is modeled through a capacity option contract. This is a contract structure that is defined by two parameters: a per-unit reservation fee $c^{res} \geq 0$ and a per-unit execution fee $c^{exe} \geq 0$. Owning q^{res} units of such contract requires a payment of $c^{res}q^{res}$ up front to the supplier, and gives the right (but not the obligation) to the buyer to receive a shipment up to the contracted capacity, q^{res}. If $q^{exe} \leq q^{res}$ units are requested, then the buyer must pay $c^{exe}q^{exe}$ to the supplier. $c^{res}q^{res}$ is the reservation cost, while $c^{exe}q^{exe}$ is the execution cost.

Clearly, this contract only differs from a traditional wholesale contract when the buyer receives new information about demand between the moment in which it sets q^{res} (capacity reservation) and the moment it requests q^{exe} (capacity execution). Thus, we assume that the buyer only knows the demand distribution before reserving capacity, while it knows the realized demand before requesting the final order.

After receiving from each supplier $i = 1, 2$ a different contract (c_i^{res}, c_i^{exe}), the buyer decides how much capacity to reserve with each supplier, q_i^{res}. After the demand is realized, the buyer request final delivery from each supplier q_i^{exe}. The sequence of buyer decisions is depicted in Figure 15.1. In the analysis below, we

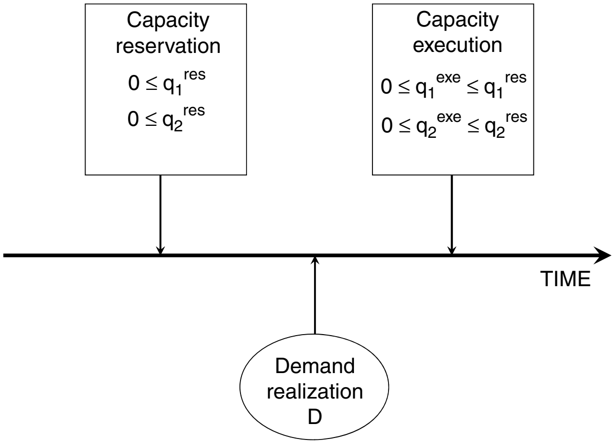


FIGURE 15.1 Sequence of events in the static model.

sort the suppliers so that $c_1^{exe} \le c_2^{exe}$. Also, to avoid trivial solutions, we assume that $c_i^{res} + c_i^{exe} < p$ for $i = 1, 2$.

We assume that the buyer is risk-neutral and maximizes its expected profit. For each realization of demand, the profit is obtained as revenue from sales $p \min\{D, q_1^{exe} + q_2^{exe}\}$, minus execution cost, $c_1^{exe} q_1^{exe} + c_2^{exe} q_2^{exe}$, minus reservation cost, $c_1^{res} q_1^{res} + c_2^{res} q_2^{res}$. Given (q_1^{res}, q_2^{res}) and a demand D, it is easy to describe the optimal execution strategy: the buyer solves:

$$C^{exe}(D, q_1^{res}, q_2^{res}) := \max_{\substack{0 \le q_1^{exe} \le q_1^{res} \\ 0 \le q_2^{exe} \le q_2^{res}}} \left\{ p \min\{D, q_1^{exe} + q_2^{exe}\} - c_1^{exe} q_1^{exe} - c_2^{exe} q_2^{exe} \right\} \tag{15.1}$$

The optimal policy is to set:

$$q_1^{exe*}(D, q_1^{res}, q_2^{res}) = \min\{D, q_1^{res}\} \text{ and } q_2^{exe*}(D, q_1^{res}, q_2^{res}) = \min\{D - q_1^{exe*}, q_2^{res}\} \tag{15.2}$$

As a result, the execution policy of the buyer has a simple greedy structure: the supplier who offers the lowest execution cost will always be used before the one with highest execution cost. Thus, the demand will be served by requesting first supply from supplier 1, and, if the capacity reserved at supplier 1 is insufficient, then the necessary additional supply will be ordered at supplier 2.

Interestingly, using equation (15.2), C^{exe} can be expressed as:

$$\begin{aligned} & (p - c_1^{exe}) \min\{D, q_1^{res}\} + (p - c_2^{exe}) \min\{(D - q_1^{res})^+, q_2^{res}\} \\ = \; & (c_2^{exe} - c_1^{exe}) \min\{D, q_1^{res}\} + (p - c_2^{exe}) \min\{D, q_1^{res} + q_2^{res}\} \end{aligned}$$

Note that $x^+ = \max\{x, 0\}$. We can thus define the buyer's capacity reservation problem as follows:

$$\max_{\substack{0 \le q_1^{res} \\ 0 \le q_2^{res}}} \left\{ \mathbb{E} C^{exe}(D, q_1^{res}, q_2^{res}) - c_1^{res} q_1^{res} - c_2^{res} q_2^{res} \right\}$$

In other words, using that $\mathbb{E} \min\{D, x\} = \int_0^x \bar{F}(s) ds$:

$$\max_{\substack{0 \le q_1^{res} \\ 0 \le q_2^{res}}} \left\{ (c_2^{exe} - c_1^{exe}) \int_0^{q_1^{res}} \bar{F}(s) ds + (p - c_2^{exe}) \int_0^{q_1^{res}+q_2^{res}} \bar{F}(s) ds - c_1^{res} q_1^{res} - c_2^{res} q_2^{res} \right\} \tag{15.3}$$

This objective is jointly concave in (q_1^{res}, q_2^{res}). We can thus find the optimal solution q_1^{res*}, q_2^{res*} to (15.3) by solving its first-order conditions. This is shown in the next theorem.

THEOREM 15.1

(i) If $\dfrac{c_1^{res} - c_2^{res}}{c_2^{exe} - c_1^{exe}} > 1$, $q_1^{res*} = 0$ *and* $q_2^{res*} = \bar{F}^{-1}\left(\dfrac{c_2^{res}}{p - c_2^{exe}}\right)$

(ii) If $\dfrac{c_1^{res} - c_2^{res}}{c_2^{exe} - c_1^{exe}} < \dfrac{c_2^{res}}{p - c_2^{exe}}$, $q_1^{res*} = \bar{F}^{-1}\left(\dfrac{c_1^{res}}{p - c_1^{exe}}\right)$ *and* $q_2^{res*} = 0$

(iii) Otherwise $q_1^{res*} = \bar{F}^{-1}\left(\dfrac{c_1^{res} - c_2^{res}}{c_2^{exe} - c_1^{exe}}\right)$ *and* $q_2^{res*} = \bar{F}^{-1}\left(\dfrac{c_2^{res}}{p - c_2^{exe}}\right) - q_1^{res*}$

Thus, the theorem identifies the optimal reservation strategy for the buyer. The optimal capacities are found using critical fractile equations. Three regions are identified. First, if $\frac{c_1^{res} - c_2^{res}}{c_2^{exe} - c_1^{exe}} > 1$ then supplier 1 should not be used at all. In fact, the condition is equivalent to $c_1^{res} + c_1^{exe} > c_2^{res} + c_2^{exe}$, i.e., supplier 1's total price is higher than supplier 2's. Hence, not only supplier 2 offers more flexibility (higher execution cost, i.e., a larger part of the payment can be postponed after demand is realized), but it is also cheaper overall. Second, if $\frac{c_1^{res} - c_2^{res}}{c_2^{exe} - c_1^{exe}} < \frac{c_2^{res}}{p - c_2^{exe}}$ then supplier 2 should not be used. This corresponds to situations where the reservation cost and/or the execution cost of supplier 2 are too high.

In any other case, both suppliers should be used at the capacity reservation stage. Both q_1^{res*} and $q_1^{res*} + q_2^{res*}$ are determined by the cost parameters and the demand distribution. The amount to reserve from supplier 2 depends on how much was reserved from supplier 1: the structure of the capacity reservation problem implies that the optimal decision is to "reserve up to" a level. This structure remains valid for $n > 2$ suppliers, as shown in Martínez-de-Albéniz and Simchi-Levi (2005). In that situation, sorting the suppliers by execution cost (i.e., $c_1^{exe} \le \ldots \le c_n^{exe} \le p$), the optimal reservation quantities $q_1^{exe*}, \ldots, q_n^{exe*}$ are determined by (when the solution is interior):

$$q_1^{res*} + \ldots + q_i^{res*} = \bar{F}^{-1}(z_i)$$

Note that for $i = 1, \ldots, n-1$, $z_i = \frac{c_i^{res} - c_{i+1}^{res}}{c_{i+1}^{exe} - c_i^{exe}}$ and $z_n = \frac{c_n^{res}}{p - c_n^{exe}}$. Note that for the solution to be interior, it is required that all suppliers must be used at a positive level. This allocation is indeed feasible when $1 \geq z_1 \geq \ldots \geq z_n$; this corresponds to having the suppliers' costs (c_i^{exe}, c_i^{res}) form a convex curve.

Furthermore, since we obtain closed-form expression for the buyer's problem, it is possible to study analytically the pricing dynamics between suppliers, as in Martínez-de-Albéniz and Simchi-Levi (2009).

In the dual-sourcing scenario (case [iii] of the theorem), the optimal reservation decision is highly sensitive to the demand distribution. Indeed, letting $z_1 = \frac{c_1^{res} - c_2^{res}}{c_2^{exe} - c_1^{exe}}$ and $z_2 = \frac{c_2^{res}}{p - c_2^{exe}} < z_1$, q_1^{res*} is equal to $\bar{F}^{-1}(z_1)$ and $q_2^{res*} = \bar{F}^{-1}(z_2) - \bar{F}^{-1}(z_1)$. In addition $\mathbb{E}_D q_1^{exe*}(D, q_1^{res*}, q_2^{res*}) = \int_0^{\bar{F}^{-1}(z_1)} \bar{F}(s)ds$ and $\mathbb{E}_D q_2^{exe*}(D, q_1^{res*}, q_2^{res*}) = \int_{\bar{F}^{-1}(z_1)}^{\bar{F}^{-1}(z_2)} \bar{F}(s)ds$. As a result, when the demand uncertainty is very low, it can be found that q_1^{res*} approaches the average demand $\bar{D}$ and q_2^{res*} approaches zero. The same occurs with the expected execution quantities. This is true because, for any family of demand distributions parameterized by the standard deviation, $\bar{F}^{-1}(z_1)$ tends to $\bar{D}$, and $\bar{F}^{-1}(z_2) - \bar{F}^{-1}(z_1)$ tends to zero as the standard deviation tends to zero. This is intuitive since in such situation no flexibility is required from the suppliers to adjust supply to demand. The buyer hence chooses the low total cost provider, supplier 1, because $c_1^{res} + c_1^{exe} \leq c_2^{res} + c_2^{exe}$. On the other hand, when demand uncertainty is very high, flexibility to adjust supply to demand becomes important. As a result, supplier 2 typically receives a larger reservation from the buyer. Indeed, as the demand standard deviation increases, $\bar{F}^{-1}(z_2) - \bar{F}^{-1}(z_1)$ increases for most parameter values (e.g., for Gamma distributions, it may decrease for a very large standard deviation; for Normal distributions, it always increases). In contrast, the capacity at supplier 1 q_1^{res*} can increase or decrease depending on the value of z_1. Typically, if $z_1 > 0.5$, q_1^{res*} decreases, while if $z_1 < 0.5$, it can increase or decrease (e.g., for Gamma distributions, it initially increases and then decreases; for Normal distributions, it increases). These dynamics are illustrated in Figure 15.2, for a Gamma distribution with different standard deviations.

Thus, while Theorem 15.1 shows that dual sourcing is optimal when $1 > \frac{c_1^{res} - c_2^{res}}{c_2^{exe} - c_1^{exe}} > \frac{c_2^{res}}{p - c_2^{exe}}$, Figure 15.2 suggests that the effective value of using the supplier portfolio increases with demand uncertainty, since both suppliers are used in larger quantities then. We can in fact calculate the value of dual sourcing compared to single sourcing by comparing the buyer's expected profit using one or both suppliers. For this purpose, define $\Pi_1 = \max_{0 \leq q_1^{res}} \left\{ \mathbb{E}C^{exe}(D, q_1^{res}, 0) - c_1^{res} q_1^{res} \right\}$, $\Pi_2 = \max_{0 \leq q_2^{res}} \left\{ \mathbb{E}C^{exe}(D, 0, q_2^{res}) - c_2^{res} q_2^{res} \right\}$ and Π_{dual} as the maximum in (15.3). Using the parameters of Figure 15.2 we compare the profits achieved under single or dual sourcing in Figure 15.3.

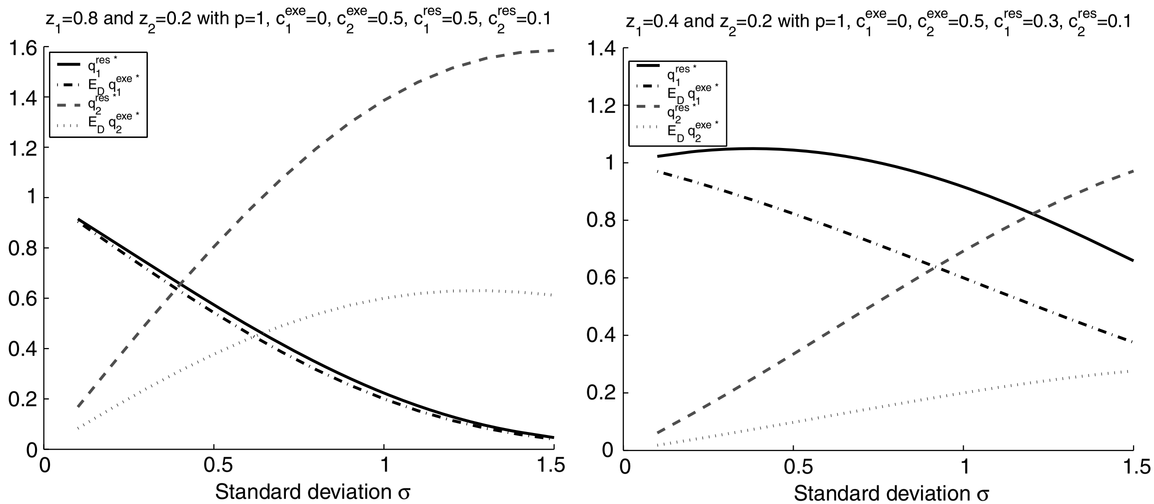


FIGURE 15.2 Reservation and expected execution quantities for different demand variabilities. In the figure, the demand is Gamma distributed with average 1 and standard deviation σ. The left-hand figure uses $z_1 = 0.8$, $z_2 = 0.2$, the right-hand figure $z_1 = 0.4$, $z_2 = 0.2$.

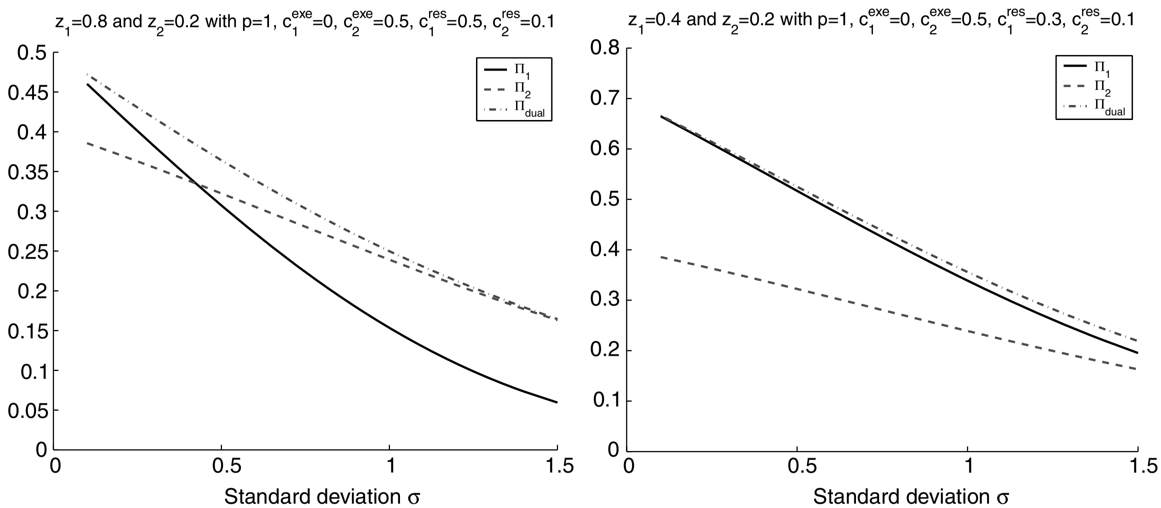


FIGURE 15.3 Comparison of Π_1, Π_2 and Π_{dual} for different demand variabilities, using the parameters of Figure 15.2.

Figure 15.3 shows that dual sourcing only brings marginal additional value when variability is low, since in that case using only supplier 1 is in fact optimal ($q_2^{res*} = 0$ when $\sigma = 0$). On the other hand, when variability is very high, dual sourcing again does not increase profit by much, since the best strategy is to use mainly supplier 2 because of its lower reservation fee. However, for intermediate values of σ, dual sourcing clearly dominates single sourcing, with profit increases in the 10% − −15% range. More generally, the insights derived from the figure hold true when (i) demand uncertainty is neither too low nor too high and (ii) the cost structure of the suppliers is sufficiently different so that dual sourcing effectively provides both lower cost and higher flexibility (which is possible when $z_2 < z_1$).

15.4 A Dynamic Model with Progressive Demand Revelation

In the previous section, we focused on a static model where demand is revealed between the reservation and the execution phase. Here we provide an alternative model to capture similar trade-offs. The contracts offered by the suppliers have now a simpler structure (wholesale contracts without optionality) but the time at which contracts are signed is different depending on the supplier. Similarly as before, the buyer sells a product at a price p per unit to serve a stochastic demand D.

Hence, we assume that the suppliers now differ in their price and lead-time. Each supplier $i = 1, 2$ offers a wholesale contract with a per-unit fee c_i. This contract must be signed L_i time units before demand is realized, since production and delivery from this supplier is not instantaneous. As a result, the buyer now needs to sequentially decide how much to procure from each of the suppliers. To avoid trivial solutions, we assume that $c_1 \leq c_2$ but also $L_1 \geq L_2$ (i.e., supplier 1 has lower cost but requires earlier commitment from the buyer), since its lead time is higher.

It is usually the case that the quality of demand forecasts improves over time. We model such a forecast evolution through a demand forecast update. This update can be obtained through a demand signal, as in Iyer and Bergen (1997), or as an advance sales realization, as in Fisher and Raman (1996). For this purpose, we let f_1 be the p.d.f. of the demand L_1 units before the demand is realized, F_1 its c.d.f. and $\bar{F}_1 = 1 - F_1$. Between that moment (L_1 time units before sales) and the next sourcing decision (L_2 time units before sales), the buyer receives a (stochastic) forecast update U_1. Given a particular update u_1, the buyer generates a new demand forecast: The demand distribution evaluated L_2 time units before the sales season is thus defined with p.d.f. $f_2(\cdot, u_1)$ and c.d.f. $F_2(\cdot, u_1)$. Because the update must be consistent with Bayes' rule, the probability of the demand being smaller than x, L_1 time units before sales (i.e., $F_1(x)$, must be equal to the expected post-update probability $\mathbb{E}_{U_1} F_2(x, U_1)$). In other words, $F_1(x) = Pr[D \leq x]$ and $F_2(x, u_1) = Pr[D \leq x | U_1 = u_1]$.

This forecast structure is commonly used in the quick-response literature. For instance, Iyer and Bergen (1997) model D as normally distributed with average U_1 and standard deviation σ. U_1 is itself normally distributed with average μ and standard deviation τ. As a result, D is normal with average μ and standard deviation $\sqrt{\sigma^2 + \tau^2}$. Hence, the demand uncertainty faced when contracting with supplier 2 is smaller than when interacting with supplier 1.

Although our results hold generally for any specification of U_1 and D, we assume an additive update structure based on Gamma distributions. Specifically, we let U_2 be Gamma distributed with average μL_2 and standard deviation $\sigma\sqrt{L_2}$; we denote g_2 and G_2 its p.d.f. and c.d.f. respectively, with $\bar{G}_2 = 1 - G_2$. We also let U_1 be Gamma distributed with average $\mu(L_1 - L_2)$ and standard deviation $\sigma\sqrt{L_1 - L_2}$, and we denote g_1 and G_1 its p.d.f. and c.d.f. respectively, with $\bar{G}_1 = 1 - G_1$. Finally, $D = U_1 + U_2$, i.e., Gamma distributed with average μL_1 and standard deviation $\sigma\sqrt{L_1}$.

This structure corresponds to a scenario where sales are generated continuously over time, at an average rate of μ and a standard deviation σ per time unit. Since the sum of such distributions turns out to be Gamma distributed, the forecast L_1 units before demand is realized is a Gamma distribution with average μL_1 and standard deviation $\sigma\sqrt{L_1}$. L_2 units before demand realization, the buyer has already received firm orders for the initial $L_1 - L_2$ time periods, U_1. At this point the new demand forecast is a Gamma distribution with average $U_1 + \mu L_2$ and standard deviation $\sigma\sqrt{L_2}$.

As a result, the buyer's sourcing decision must balance two elements. On the one hand, using supplier 1 results in a lower per-unit cost compared to procurement from supplier 2. On the other hand, delaying the procurement decision until as late as possible (supplier 2) is beneficial since the risk of a supply/demand mismatch is lower. Denoting q_1^{dyn}, q_2^{dyn} the quantities purchased at each supplier, the sequence of buyer decisions is depicted in Figure 15.4.

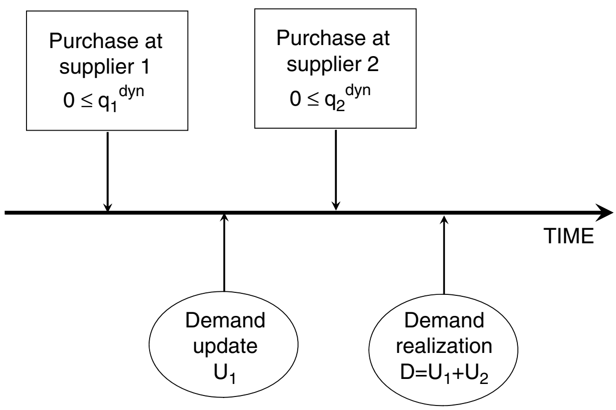


FIGURE 15.4 Sequence of events in the dynamic model.

Assuming that the buyer is risk-neutral, its objective is to solve:

$$\max_{0\leq q_1^{dyn}} \left\{\mathbb{E}_{U_1} M(q_1^{dyn}, U_1) - c_1 q_1^{dyn}\right\} \tag{15.4}$$

where

$$M(q_1^{dyn}, u_1) := \max_{0\leq q_2^{dyn}} \left\{p\mathbb{E}_{U_2} \min\{u_1 + U_2, q_1^{dyn} + q_2^{dyn}\} - c_2 q_2^{dyn}\right\} \tag{15.5}$$

This is in fact a standard two-period inventory problem with correlated demand structure. Since there are no fixed costs associated with an order, the optimal policy is to follow a state-dependent base-stock policy. Namely, there exists $b_2(u_1)$ such that $q_2^{dyn*} = \max\{0, b_2(u_1) - q_1^{dyn}\}$; and b_1 such that $q_1^{dyn*} = b_1$ (if there was an initial inventory level q_0, the optimal order would be $q_1^{dyn*} = \max\{0, b_1 - q_0\}$). In fact, using the additive demand structure, the base-stock levels can be found in closed form, as stated in the next theorem.

Theorem 15.2 *The order quantity from each supplier is uniquely characterized by:*

- $q_2^{dyn*} = \max\{0, b_2 + U_1 - q_1^{dyn*}\}$ *where* $b_2 = \bar{G}_2^{-1}\left(\frac{c_2}{p}\right)$*; and*
- q_1^{dyn*} *satisfying*

$$\frac{c_2 - c_1}{p} = \mathbb{E}_{U_1} \max\left\{0, \bar{G}_2(b_2) - \bar{G}_2(q_1^{dyn*} - U_1)\right\}.$$

Theorem 15.2 implies that the optimal quantity purchased at the suppliers is calculated through critical fractile equations. In this case, in contrast with Theorem 15.1, the critical fractile levels are $\frac{c_2-c_1}{p}$ and $\frac{c_2}{p}$. The impact of lead time is captured in the shape of the distributions of U_1, U_2. Furthermore, the result can immediately be extended to $n > 2$ suppliers. Assuming that $c_1 \leq \ldots \leq c_n$ and $L_1 \geq \ldots \geq L_n$, and keeping an additive demand structure $D = U_1 + \ldots + U_n$, where for $i = 1, \ldots, n$, U_i is Gamma distributed with average $\mu(L_i - L_{i+1})$ and standard deviation $\sigma\sqrt{L_i - L_{i+1}}$ (letting $L_{n+1} = 0$), the buyer should purchase q_i^{dyn*} from each supplier according to a state-dependent base-stock policy where the optimal base-stock level is $b_i + (U_1 + \ldots + U_{i-1})$ for a well-chosen parameter b_i, which can be interpreted as a safety stock. b_i depends exclusively on $\frac{c_{i+1}-c_i}{p}, \frac{c_{i+2}-c_{i+1}}{p}, \ldots, \frac{c_n-c_{n-1}}{p}, \frac{c_n}{p}$.

In addition, while the structure of the problem is well-known, and the optimal policy is a base-stock policy as one would expect, the theorem in fact reveals how the cost and lead-times of the suppliers influence the sourcing decision. First, the quantity q_2^{dyn*} is in fact stochastic, as it depends on the realization of the forecast update U_1 (one can compute its expected value: $\mathbb{E}_{U_1} q_2^{dyn*} = \int_{q_1^{dyn*}-b_2}^{\infty} \bar{G}_1(s)ds$). This quantity is decreasing in c_2 and increasing in c_1. Similarly, q_1^{dyn*} decreases in c_1 and increases in c_2. This implies that the introduction of a second, shorter lead-time supplier, reduces q_1^{dyn*}. Indeed, if $c_2 < \infty$ the buyer expects to use supplier 2 when the update is high enough, while it would never do so if $c_2 = \infty$. Second, the effect of lead-time is captured in the distribution of U_1 and U_2. Keeping L_1, c_1, c_2 constant, increasing L_2 implies that supplier 2 requires earlier commitment. As a result, q_1^{dyn*} increases, and q_2^{dyn*} decreases. Figure 15.5 illustrates this effect.

The impact of demand uncertainty on the optimal order quantities also provides interesting insights. The effect of a change in the standard deviation of U_1 depends on the shape of the distribution of U_2. Indeed, if $\bar{G}_2(x)$ is concave for $x \leq b_2$ (e.g., b_2 is lower than the mode of U_2), then increasing the standard deviation of U_1 (and keeping its mean constant) leads to a higher $\mathbb{E}_{U_1} \max\{0, \bar{G}_2(b_2) - \bar{G}_2(q_1^{dyn} - U_1)\}$ for any q_1^{dyn}. This results in a lower q_1^{dyn*}, and a stochastically higher q_2^{dyn*}. This is intuitive, because with more uncertain U_1 it is better to avoid the uncertainty by buying more at supplier 2 with knowledge of the forecast update. However, when $\bar{G}_2(x)$ is convex, q_1^{dyn*} may increase or decrease. The effect of a change in the standard deviation of U_2 is more complex. On the one hand, it affects $\bar{G}_2(q_1^{dyn} - U_1)$ directly, and depending on the value of $\frac{c_2-c_1}{p}$, q_1^{dyn*} may increase or decrease. On the other hand, it affects b_2, which may increase or decrease depending on the value of $\frac{c_2}{p}$. Overall, increasing the standard deviation of U_2 may increase or decrease q_1^{dyn*} and q_2^{dyn*} depending on the parameters.

Furthermore, we can again evaluate the advantages of using both suppliers compared to single sourcing in the dynamic model. For this purpose, we define $\Pi_1 = \max_{0 \leq q_1^{dyn}} \{p\mathbb{E}_D \min\{D, q_1^{dyn}\} - c_1 q_1^{dyn}\}$, $\Pi_2 = \mathbb{E}_{U_1} M(0, U_1)$ and Π_{dual} as the maximum in (15.4). Similarly as with the static model, using the parameters of Figure 15.5 we compare the profits achieved under single or dual sourcing in Figure 15.6.

We observe in the figures that when L_2 is close to zero or close to $L_1 = 1$, then planning to use dual sourcing only brings marginal value beyond single sourcing. For intermediate values of L_2, on the other hand, the advantage is more significant.

Finally, it is worth mentioning that the dynamic model and the static model can in fact overlap and capture identical situations. Consider the dynamic model with $L_2 = 0$ (i.e., the second supplier can produce after observing realized demand). In this case, $q_2^{dyn*} = (D - q_1^{dyn*})^+$ $(b_2 = 0)$, and $\frac{c_1}{c_2} = \mathbb{P}\{D \geq q_1^{dyn*}\}$. On the other hand, consider the static model with $c_1^{res} = c_1$, $c_1^{exe} = 0$, $c_2^{res} = 0$ and $c_2^{exe} = c_2$. As a result, $q_2^{res*} = \infty$, $q_1^{exe*} = \min\{D, q_1^{res*}\}$ and $q_2^{exe*} = (D - q_1^{res*})^+$. q_1^{res*} is such that $\mathbb{P}\{D \geq q_1^{res*}\} = z_1 = \dfrac{c_1}{c_2}$. Hence, $q_1^{dyn*} = q_1^{res*}$ and $q_2^{dyn*} = q_2^{exe*}$. This shows that both models capture similar aspects of flexibility, namely, the

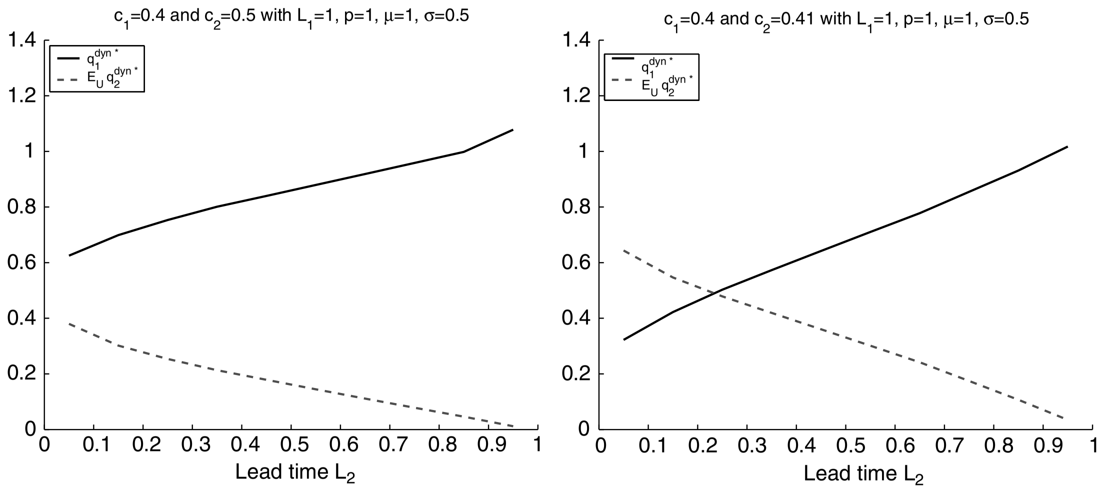


FIGURE 15.5 **Sourcing quantities q_1^{dyn*} and $E_{U_1} q_2^{dyn*}$ as function of L_2, keeping $L_1 = 1$ constant. The demand parameters are $\mu = 1$, $\sigma = 0.5$. The left-hand figure uses $c_1 = 0.4$, $c_2 = 0.5$, the right-hand figure $c_1 = 0.4$, $c_2 = 0.41$; $p = 1$ in both cases.**

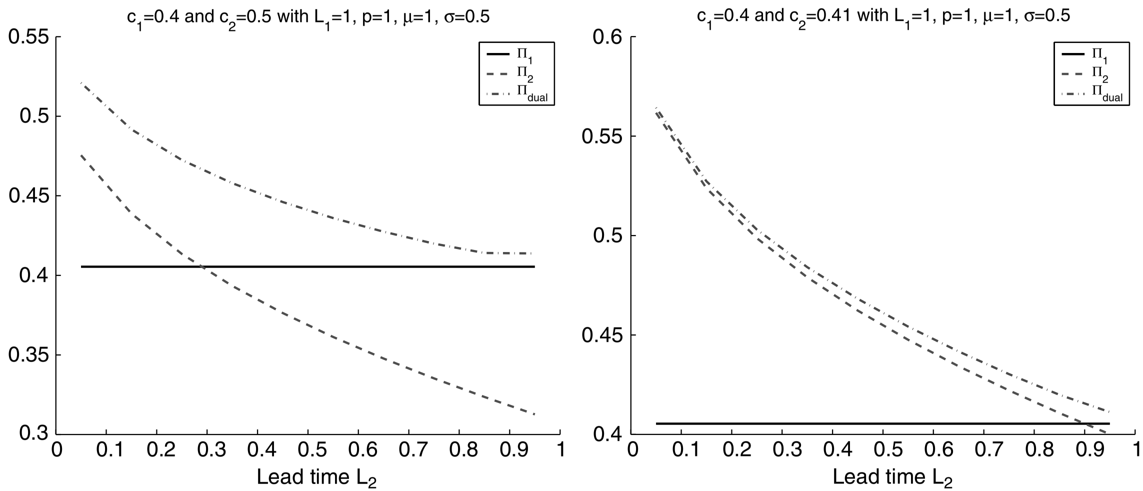


FIGURE 15.6 Comparison of Π_1, Π_2 and Π_{dual} for different lead-times L_2, given $L_1 = 1$, using the parameters of Figure 15.5.

trade-off between paying a higher/lower cost of production with less/more information about demand.

15.5 Conclusions

The models presented in this chapter analyze the basic trade-off between cost and flexibility in procurement. The static model focuses on a standard stochastic demand model where a production decision must be taken before and after demand is realized (reservation and execution). Flexibility is modeled through different supplier cost structures. In contrast, the dynamic model uses a more sophisticated demand model, with forecasts updates, where production decisions are taken after each update. Flexibility is captured through the lead times of the suppliers, where a shorter lead time allows to use more accurate demand information when placing the order. In the extreme case, when the lead time of a supplier is zero, the decision can be taken after demand is known.

The value of using a portfolio is created from using the lowest cost supplier for the demand that materializes with high likelihood, and using the higher cost but more flexible supplier to fulfill the fraction of demand that has higher uncertainty. Through both models, we demonstrate that using dual sourcing clearly dominates single sourcing, especially when the improvement in demand information between the first and second production decisions is neither too high nor too low.

While the exposition focused on the case of two suppliers, the analysis can be extended to $n > 2$ suppliers in a straightforward way. Other features can be added to both models without requiring any different approach: holding costs and lost sales penalties can be added; uncertainty in supplier costs or reliability can be incorporated; using a multiple period demand model is also simple (for the static model and for the dynamic model with $L_1 - L_2 = 1$, see Fukuda (1964)). This is possible since the buyer's cost is jointly convex in all sourcing decisions. As a result, these optimal decisions can be characterized through first-order optimality conditions, which are expressed through critical fractile equations. Some extensions however lead to a different model structure. Indeed, no set-up costs are included in the model. Including these would require adding a binary decision (whether to use or not use a supplier) in the buyer's decision. This changes the nature of the optimization problem. For example, in the dynamic model, if a set-up cost of K must be payed when using a supplier, then the optimal sourcing decision would have a (s, S) structure, see Sethi and Cheng (1997). More generally, if the suppliers' costs exhibit economies of scale, similar changes in the structure of the problem occur.

The use of multiple sourcing as a mechanism to hedge demand risks opens a number of interesting research questions. First, the focus of this chapter has been on using the supply portfolio to adjust supply to demand. In many practical settings, this strategy can be complemented with *demand management* decisions, such as price adjustments or marketing campaigns to reduce or increase demand. These two activities ultimately have the same objective: to minimize the supply/demand mismatch. How to effectively combine them is an intriguing question, and some advances on the matter have been provided in Fu et al. (2007). Second, the approach offered here assumes that the buyer is risk-neutral. However, in industries where demand uncertainty is high, low demand realizations for several seasons can lead companies to bankruptcy. The toy industry is a good example of this, see Lago (2007). As a result, firms are sensitive to risk and need to consider not only the expected profits derived from multiple sourcing, but also higher moments. It is hence important to provide managers with methods and tools *to control risk*, so that they can choose the appropriate level of risk and determine the strategy that maximizes expected profit within the risk tolerance. This leads to an efficient frontier related for the supply strategy, see Martínez-de-Albéniz and Simchi-Levi (2006). Other approaches inspired from finance can be used as well, such as the Supply-At-Risk of Haksöz and Kadam (2009). In any case, describing the risk of a supply portfolio strategy is an important requirement for practical applications, and some further research on this line is expected. Third, while here supplier costs are exogenously given, the use of dual sourcing generally affects *supplier incentives*. In a competitive environment, the use of dual sourcing has strong implications, as observed in Martínez-de-Albéniz and Simchi-Levi (2009). For example, companies such as Toyota actively manage their supplier portfolios over time to drive competition between suppliers, and at the same time derive the benefits of a stable supply base. Furthermore, in a supplier development context, the improvement efforts between suppliers are very different depending on its role in the supplier portfolio. Indeed, the order placed with the lowest cost supplier tends to be stable, while the order placed at the flexible supplier has very high variability, much higher than the demand itself. As a result, the flexible supplier improvement efforts should be directed to accommodating order variability better (at a lower cost), while the lowest cost supplier should focus on reducing absolute costs even more. Studying such incentive dynamics is an interesting direction for future work.

REFERENCES

Anupindi R., Akella R. 1993. Diversification under supply uncertainty. *Management Science*, **39**(8):944–963.

Babich V., Burnetas A.N., Ritchken P.H. 2007. Competition and diversification effects in Supply Chains with Supplier Default Risk. *Manufacturing and Service Operations Management*, **9**(2):123–146.

Calvo E., Martínez-de-Albéniz V. 2010. Sourcing Strategies for Short Life-Cycle Goods. Working paper. IESE Business School, Barcelona, Spain.

Chaturvedi A., Martínez-de-Albéniz V. 2008. Optimal procurement design in the presence of supply risk. *Manufacturing & Service Operations Management*, **13**(2):227–243.

Chaturvedi A., Martínez-de-Albéniz V. 2009. Safety stock or excess capacity: Trade-offs under Supply Risk. Working paper. IESE Business School, Barcelona, Spain.

Dada M., Petruzzi N.C., Schwarz L.B. 2007. A newsvendor's procurement problem when suppliers are unreliable. *Manufacturing and Service Operations Management*, **9**(1):9–32.

Federgruen A., Yang N. 2008. Selecting and portfolio of suppliers under demand and supply risks. *Operations Research*, **56**(4):916–936.

Feng Q., Sethi S.P., Yan H., Zhang H. 2005. Are base-stock policies optimal in inventory problems with multiple delivery modes? *Operations Research*, **54**(4):801–807.

Fisher M.L., Raman A. 1996. Reducing the cost of demand uncertainty through accurate response to early sales. *Operations Research*, **44**(1):87–99.

Fu Q., Lee C.-Y., Teo C.-P. 2010. Procurement management using option contracts: Random spot price and the portfolio effect. *IIE Transactions*, **42**(11): 793–811.

Fu Q., Zhou S.X., Chao X., Lee C.-Y. 2007. Combined pricing and portfolio option procurement. Working paper. Hong Kong: Hong Kong University of Science and Technology.

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

Gerchak Y., Parlar M. 1990. Yield randomness/cost tradeoffs and diversification in the EOQ model *Naval Research Logistics*, **37**(3):341–354.

Haksöz Ç., Kadam A. 2009. Supply portfolio risk. *The Journal of Operational Risk*, **4**(1):59–77.

Iyer A.V., Bergen M.E. 1997. Quick response in manufacturer-retailer channels. *Management Science*, **43**(4):559–570.

Lago A. 2007. Famosa global production strategy. Case study P-1085-E. IESE Publishing, Barcelona, Spain.

Martínez-de-Albéniz V. 2005. Pricing in a duopoly with a lead time advantage. Working paper. IESE, Barcelona, Spain.

Martínez-de-Albéniz V., Simchi-Levi D. 2005. A portfolio approach to procurement contracts. *Production and Operations Management*, **14**(1):90–114.

Martínez-de-Albéniz V., Simchi-Levi D. 2006. Mean-variance trade-offs in supply contracts. *Naval Research Logistics*, **53**(7):603–616.

Martínez-de-Albéniz V., Simchi-Levi D. 2009. Competition in the supply option market. *Operations Research*, **57**(5):1082–1097.

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

Moinzadeh K., Nahmias S. 1988. A continuous review model for an inventory system with two supply modes. *Management Science*, **34**(6):761–773.

Moinzadeh K., Schmidt C.P. 1991. An $(S-1, S)$ inventory system with emergency orders. *Operations Research*, **32**(2):308–321.

Nagali V. et al. 2008. Procurement risk management (PRM) at Hewlett-Packard Company. *Interfaces*, **38**(1):51–60.

Parlar M., Wang D. 1993. Diversification under yield randomness in inventory models. *European Journal of Operational Research*, **66**(1):52–64.

Sethi S.P., Cheng F. 1997. Optimality of (s, S) policies in inventory models with Markovian demand. *Operations Research*, **45**(6):931–939.

Song J.-S., Zipkin P.H. 2007. Newsvendor Problems with Sequentially Revealed Demand Information. Working paper, Duke University.

Tagaras G., Vlachos D. 2001. A periodic review inventory system with emergency replenishments. *Management Science*, **47**(3):415–429.

Tomlin B. 2006. On the value of mitigation and contingency strategies for managing supply chain disruption risks. *Management Science*, **52**(5):639–657.

Veeraraghavan S., Scheller-Wolf A. 2008. Now or later: A simple policy for effective dual sourcing in capacitated systems. *Operations Research*, **56**(4):850–864.

Wan Z., Beil D.R. 2008. Bargaining power and supply base diversification. Working paper. Ann Arbur, Michigan: University of Michigan.

Whittemore A.S., Saunders S.C. 1977. Optimal inventory under stochastic demand with two supply options. *SIAM Journal of Applied Mathematics*, **32**(2):293–305.

Wu D.J., Kleindorfer P.R. 2005. Competitive options, supply contracting, and electronic markets. *Management Science*, **51**(3):452–466.

Wu D.J., Kleindorfer P.R., Zhang J.E. 2002. Optimal bidding and contracting strategies for capital-intensive goods. *European Journal of Operations Research*, **137**(3):657–676.

CHAPTER SIXTEEN

An Opportunity Cost View of Base-Stock Optimality for the Warehouse Problem

NICOLA SECOMANDI

This work considers the so called warehouse problem, which is a prototypical problem of the trading activity of a merchant in a commodity market. It is known that the merchant's optimal trading policy for this problem has a base-stock structure. The exiting proofs of this result hinge on marginal analysis, and may not be easily accessible to managers. This work provides an elementary derivation of the optimality of this structure relying almost exclusively on geometric arguments based on the notion of opportunity cost of a trade, a concept familiar to commodity merchants. Some aspects of managerial relevance associated with this structure are also discussed. It is hoped that the material presented in this work would be of interest to managers involved in the merchant management of commodity storage.

Acknowledgments: I thank three reviewers for their useful feedback.

Handbook of Integrated Risk Management in Global Supply Chains, First Edition. Edited by Panos Kouvelis, Lingxiu Dong, Onur Boyabatli, and Rong Li.

16.1 Introduction

Commodity storage plays a fundamental role in commodity industries (Williams and Wright 1991).The so-called warehouse problem is a foundational model of the merchant management of a commodity storage facility. More than 60 years ago, Cahn (1948) succinctly introduced this problem in the literature as follows: "Given a warehouse with fixed capacity and an initial stock of a certain product, which is subject to known seasonal price and cost variations, what is the optimal pattern of purchasing (or production), storage and sales?"

In this description, Cahn uses fixed capacity to refer to limited space availability. Thus, this version of the problem is best described as the uncapacitated case, that is, without flow rate limits. In contrast, the capacitated version of the warehouse problem refers to the situation where the warehouse features both space and flow rate constraints. This version of the problem arises in applications such as the merchant management of natural gas storage (Maragos 2002, Geman 2005, Lai et al. 2010).

In both the uncapacitated and the capacitated cases the optimal policy for the warehouse problem is of the base-stock type. That is, there exist two critical inventory levels, the base-stock targets, such that, during a given time period and for a given spot price, it is optimal to purchase or sell inventory to bring the inventory level as close as possible to these levels when the inventory level is respectively below or above the lower and upper base-stock targets; it is optimal to store otherwise (in the uncapacitated case the base-stock targets are reachable from every inventory level). In the uncapacitated case this has been shown by Bellman (1956), Dreyfus (1957), and Charnes et al. (1966); in the capacitated case this has been established, independently, by Rempala (1994) in a special case, and by Secomandi (2010) in a more general case.

The base-stock structure is insightful and characterizes the inventory trading policy computed by methods employed in practice, such as that described by Maragos (2002) and studied by Lai et al. (2010). The current literature uses marginal analysis to establish the optimality of this structure. The basic argument behind this type of analysis is the comparison of the marginal cost/revenue associated with an immediate trade with the (expected) discounted future marginal value associated with the resulting inventory level, a basic insight that goes back to Massé (1946), as cited by Wallace and Fleten (2003). Although insightful, this literature may not be easily accessible to managers.

The premise of this work is that it is important for managers to understand why the basestock target structure is optimal. This is true whether they are already using methods whose optimal policy has this structure, perhaps without explicitly realizing it, or if they are currently using methods that are not consistent with this structure. Consequently, the goal of this work is to provide a simple derivation of the optimality of the base-stock structure almost entirely based on intuitive geometric arguments that rely on the notion of opportunity cost of traded inventory. The familiarity of commodity traders with the concept of opportunity cost, even if only qualitatively, should make the analysis of this work accessible to them.

It is important to emphasize that the analyses available in the literature also rely on a notion of opportunity cost. But the key difference between these analyses and that conducted here is that the latter uses the concept of opportunity cost of a decision, which one could label as *total* opportunity cost of this decision, while the existing analyses use the concept of *marginal* opportunity cost. Although these are related concepts, the resulting analyses are different.

The remainder of this work proceeds as follows. Section 16.2 provides a simple motivating example that illustrates the base-stock structure of the optimal policy. Section 16.3 presents a formal model of the warehouse problem. Section 16.4 establishes the structure of the optimal policy based on the stated opportunity cost argument. Section 16.5 discusses some managerial aspects related to this structure. Section 16.6 concludes.

16.2 A Simple Motivating Example

Consider the following simple example discussed in Secomandi (2010). Suppose you are a merchant operating in the wholesale market for a given commodity. For the next three months you have access to a warehouse where this commodity can be stored. This facility has a space limit, which is normalized to 1 unit, that is, you can only store a finite amount of commodity, but otherwise one could fill up or empty the facility in a single month.

At the beginning of the first month you can decide to purchase the commodity from the spot market and inject it into the facility during this month. At the beginning of the second month you can purchase additional commodity from this market and inject it into the facility, you can withdraw some of the commodity that you have injected in the previous month and sell it into the wholesale market, or continue to store the commodity that is already in the facility. In the last month you withdraw and sell any amount of commodity available in storage.

To avoid clutter, suppose that the injection and withdrawal marginal costs and the inventory holding cost are zero, and let the discount factor be 1. Furthermore, suppose that the price dynamics during the next three months are known to be "medium," "low," and "high," that is, there is no uncertainty in these prices (e.g., these could represent futures prices for the next three months as of the first month).

Starting with an empty warehouse, it is optimal to fill up the facility in the second month at the low price, and sell the entire inventory in the third month at the high price. This is a base-stock policy with month dependent buy-and-inject (BI) and withdraw-and-sell (WS) base-stock levels, specified as follows: In month 1 the BI base-stock level is equal to 0 (the minimum space) and the WS base-stock level is equal to 1 (the maximum space); in month 2 both the BI and WS base-stock levels are equal to 1; and in month 3 they are both equal to 0. Interestingly, the base-stock levels are rather simple, being either equal to 0 or to 1.

But suppose now that while it is still possible to empty the facility in a single month, filling it up requires more than one month, but less than two months.

For concreteness, suppose that you can fill up 2/3 of the facility in a single month. You will see that this has a significant impact on the optimal base-stock levels.

You still want to have a full facility at the beginning of month 3, so that you can withdraw and sell all the available inventory at the high price during this month. You would also like to purchase and inject as much inventory as possible during month 2, at the low price. But given the limit on the injection capacity, you are now forced to purchase and inject some commodity in month 1, at the medium price. Clearly, in month 1 you would limit yourself to purchase and inject the minimum amount of commodity necessary, which is 1/3, and would purchase and inject the remaining 2/3 during month 2.

This is also a base-stock policy: In month 1 both the BI and WS basestock targets are equal to 1/3; in month 2 both these targets are equal to 1 (notice that the BI base-stock target in this month is not reachable from inventory levels below 1/3 because the injection capacity is 2/3); and in month 3 both these targets are equal to 0. Thus, the effect of limited injection capacity has substantially changed the optimal trading policy. It remains of the base-stock type, but one base-stock target may not be reachable from some inventory levels, and the base-stock targets are no longer always equal to either 0 or 1.

These examples are simple and do not require a formal analysis. But to gain a deeper understanding of why a base-stock target policy is optimal, the first step is to formulate a model.

16.3 Model

This section formulates a simplified version of the model presented by Secomandi (2010). This is a finite horizon periodic review dynamic programming model, where inventory trading decisions are made at the beginning of each of a finite number of time periods, each of equal length. It is useful to think of these time periods as the months corresponding to the delivery periods following the maturities of futures contracts on the commodity that is being traded (see Luenberger [1998, Chapter 10] for the definition of futures contract).

The main simplification relative to the model of Secomandi (2010) is that here the price dynamics are assumed to be deterministic. These prices can be interpreted as the discounted futures prices that are available at the beginning of the time horizon for each relevant maturity. Thus, there is no need to model the discount factor, which is thus taken to be 1.

These assumptions greatly simplify the exposition, but, as discussed in Section 16.5, some heuristic models used in practice are based on periodic reoptimizations of a deterministic model of the type described below. Moreover, when prices evolve stochastically, the base-stock structure depends on their realizations at each decision epoch, in addition to time.

To avoid clutter, there is also no physical inventory holding cost (the model of Secomandi (2010) allows for a proportional holding cost to be incurred in each stage). Set $\mathcal{J}$, defined as $\{1, \ldots, J\}$, where J is an integer greater than 0, is used to index the time of each inventory trading and operational decision; that is, the "jth" decision, $j \in \mathcal{J}$, is made at the start of time period j. An inventory trading and operational decision is referred to as an action and is denoted by a.

A positive action corresponds to a purchase followed by an injection, a negative action to a withdrawal followed by a sale, and zero is the do nothing (DN) action. The trading part of an action is performed at the beginning of each stage, when its payoff is also accounted for. The operational part of an action, that is, an injection or a withdrawal, is executed during a stage, immediately after its associated purchase or sale at the beginning of this stage. This means that inventory injected/withdrawn sold in a given stage is available/unavailable in storage in the next stage.

Let $(\hat{s}_1, \ldots, \hat{s}_J)$ be the vector of relevant prices for stages 1 through J (recall that these can be interpreted as discounted futures prices as of stage 1). There are positive marginal costs for withdrawing and injecting the commodity. It is useful to define as s_j the marginal cost of buying *and* injecting one unit of commodity, that is, this is $\hat{s}_j$ plus the marginal injection cost. By letting c denote the sum of the marginal withdrawal and injection costs, the net price of withdrawing and selling one unit of commodity is $s_j - c$. Thus, the immediate payoff of action a in stage $j \in \mathcal{J}$ is:

$$p_j(a) := \begin{cases} -(s_j - c)a, & a < 0, \\ 0, & a = 0, \quad \forall j \in \mathcal{J} \\ -s_j a, & a > 0, \end{cases}$$

In each stage, this function is piecewise linear and concave: it has a kink at zero, where its slope decreases from $-(s_j - c)$ to $-s_j$.

The warehouse minimum and maximum inventory levels are normalized to 0 and 1, respectively, so that the feasible inventory set is $\mathcal{X} := [0, 1]$. There are limits to the amount of inventory that can be injected or withdrawn during a stage. These are the injection and withdrawal capacities $\overline{C} > 0$ and $\underline{C} < 0$, respectively, which are constant across stages. It is assumed that both $\overline{C}$ and $-\underline{C}$ are no larger than 1. In the uncapacitated case both $\overline{C}$ and $-\underline{C}$ are equal to 1. In the capacitated case at least one of these quantities is strictly less than 1; in the example discussed in Section 16.2, in the capacitated case $\overline{C} = 2/3$.

An optimal inventory trading policy for a merchant that controls the warehouse can be obtained by solving the finite horizon dynamic program (16.1)–(16.2) formulated below. Set $\mathcal{J}$ indexes the stages and the state space in stage j is $\mathcal{X}$. Denote by $V_j(x)$ the optimal value function in stage j and state x, which is the sum of all the cash flows accumulated from stage j through stage J under an optimal policy. The dynamic programming recursion is:

$$V_{J+1}(x) := 0, \ \forall x \in \mathcal{X} \tag{16.1}$$

$$V_j(x) = \max_{a \in [\underline{C} \vee (-x), \overline{C} \wedge (1-x)]} p_j(a) + V_{j+1}(x + a), \ \forall j \in \mathcal{J}, \ x \in \mathcal{X} \tag{16.2}$$

Expression (16.1) sets boundary conditions. Recursion (16.2) links the value function across stages and states through the determination of an optimal action in each stage and state. In stage j and state x, this optimization entails maximizing the sum of the immediate payoff corresponding to feasible action a, $p_j(a)$, and the value of the resulting inventory level in the next stage, $V_{j+1}(x+a)$; notice that $\cdot \wedge \cdot \equiv \min\{\cdot,\cdot\}$ and $\cdot \vee \cdot \equiv \max\{\cdot,\cdot\}$. Moreover, in each stage and state all the feasible actions in this state are available, but at most one of them is allowed to be chosen optimally. This can be shown to be without loss of generality.

By introducing separate decision variables for the BI and WS actions, it is also possible to formulate model (16.1)–(16.2) as a linear program, which is straightforward to solve to optimality with any of the available linear programming solvers. However, the dynamic programming formulation facilitates the analysis of the structure of the optimal policy.

16.4 Base-Stock Optimality

To illustrate the ensuing results, this section uses a slight variation of the example discussed in Section 16.2 with the following parameter values: $c = 0.40$, $\underline{C} = -1$, $\overline{C} = 2/3$, $J = 3$, $s_1 = 2.20$, $s_2 = 1.20$, and $s_3 = 4.20$. The main difference between this example and the capacitated version of that discussed in Section 16.2 is the presence of nonzero marginal injection and withdrawal costs, reflected in the positive value of c (these marginal costs are both equal to 0.20). The specific values for the prices in each stage are chosen so that the figures illustrated below clearly display what they are intended to convey. However, these prices may not be realistic.

It is useful to begin the analysis of the structure of the optimal policy by pointing out an important property of the optimal value function: in each stage j the function $V_j(x)$ is piecewise linear and concave (Secomandi 2010). This function could be linear (weakly concave) in some stage; e.g., $V_J(x) = 0$ if $s_J < c$, and $V_J(x) = (s_J - c)x$ if $s_J > c$. But in general it is a strictly concave function of inventory, which means that its slope decreases at each of the breakpoints that define it as inventory increases.

Figure 16.1 illustrates the optimal value function in each of the three stages in the illustrative example. The optimal value function is a straight line in stage 3, and it has a kink at inventory level 1/3 in both stages 1 and 2. Specifically, this function in these stages is as follows:

$$
\begin{aligned}
V_3(x) &= 3.8x, \\
V_2(x) &= \begin{cases} 5.2/3 + 3.8x, & x \in [0, 1/3] \\ 2.6 + 1.2x, & x \in [1/3, 1] \end{cases} \\
V_1(x) &= \begin{cases} 6.8/3 + 2.2x, & x \in [0, 1/3] \\ 2.4 + 1.8x, & x \in [1/3, 1] \end{cases}
\end{aligned}
$$

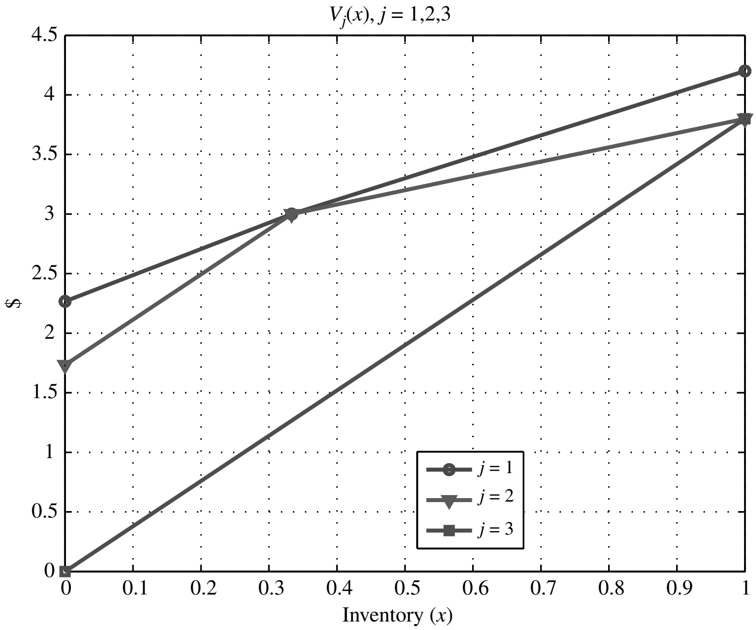


FIGURE 16.1 **The optimal value function in each stage in the illustrative example.**

In each stage $j \in \mathcal{J}$ and state $x \in \mathcal{X}$, the opportunity cost of a feasible action is the following function, which depends on both the inventory level x and the feasible action a:

$$OC_j(a, x) := V_{j+1}(x) - V_{j+1}(x + a) \tag{16.3}$$

The opportunity cost of a feasible WS action in a given stage and state is the total future value foregone by reducing the available inventory by the corresponding amount. In stage j and state x, this value is naturally measured as the difference on the right hand side of (16.3), that is, the difference between the total future values of entering stage $j + 1$ with x and $x + a < x$ units of inventory, respectively.

For a feasible BI action, the concept of opportunity revenue would be more pertinent, (i.e., the total future value created by increasing the available inventory by a given amount in a given stage and state). In stage j with inventory level x, the opportunity revenue of feasible BI action a would be $V_{j+1}(x + a) - V_{j+1}(x)$. But because the opportunity revenue of such an action is identical to the negative of its opportunity cost, it is sufficient to introduce only the opportunity cost notation.

The opportunity cost of the DN action is identically zero in every stage and state.

The notion of opportunity cost is useful to define the concept of economic action. In stage j with inventory level x, a feasible action a is said to be economic if its immediate payoff is equal to or exceeds its opportunity cost: $p_j(a) \geq OC_j(a, x)$. This definition is intuitive: it simply reflects the fact that it is optimal to perform a feasible action *only if* its immediate payoff is equal to or exceeds its opportunity cost. In other words, if an action is not economic then it is suboptimal.

According to the definition of economic action, the DN action is economic at every feasible inventory level, because in every stage j it holds that $p_j(0) = 0$ and $OC_j(0, x) \equiv 0$ for every feasible inventory level x. Thus, it is useful to interpret the DN action as being both a degenerate WS action and a degenerate BI action; in contrast a negative (respectively, positive) action is referred to as a proper WS (respectively, BI) action.

The following analysis proceeds to show that in each stage the set of feasible inventory levels is subdivided into three regions (subsets): one where only BI actions (both proper and degenerate) are economic; one where only the DN action is economic, that is, no proper BI and WS actions are economic; and one where only WS actions (both proper and degenerate) are economic, respectively. This then implies that in each stage only one type of action is respectively optimal in each of these three subsets. More formally, there exist no more than two stage dependent optimal base-stock targets $\underline{b}_j \leq \overline{b}_j$ that subdivide the set of feasible inventory levels $\mathcal{X}$ into the three sets $[0, \underline{b}_j]$, $[\underline{b}_j, \overline{b}_j]$, and $[\overline{b}_j, 1]$, where only BI, DN, and WS actions are respectively economic, and hence optimal.

Proceeding in steps, this is now argued geometrically using the notion of opportunity cost in all but the last steps. Figure 16.2 will be used to illustrate the analysis.

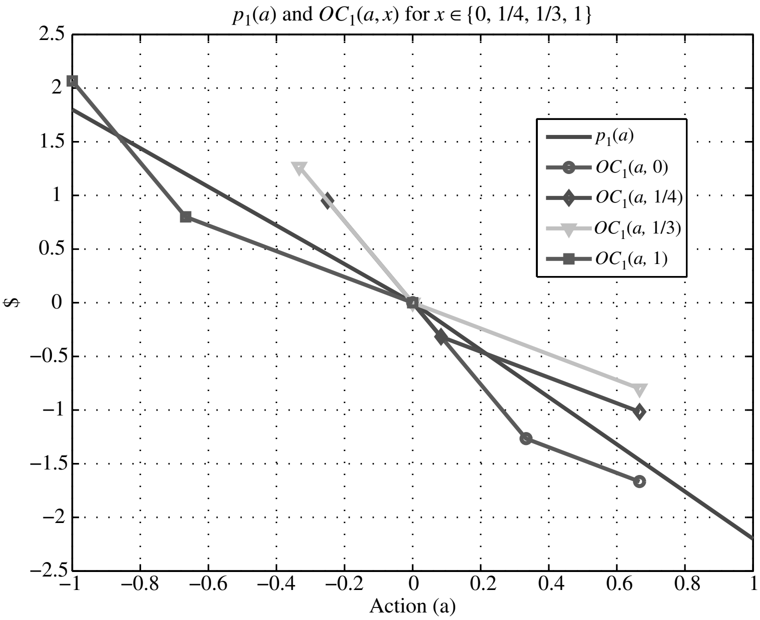


FIGURE 16.2 The immediate payoff and the opportunity cost in stage 1 at different inventory levels.

Step 1: Monotonicity of economic actions for a given inventory level. Given inventory level x in stage j, consider the opportunity cost as a function of a feasible action a. Since the function $V_j(x)$ is concave in x, it follows from the definition of opportunity cost that the function $OC_j(a, x)$ is convex in a given x; also recall that it equals zero at $a = 0$. Combining these observations with the fact that the immediate payoff is a concave and continuous function of the action, zero at zero, and decreasing when the action is positive, it is easy to see that the following monotonicity property of economic actions holds: If an action is economic at a given inventory level, then an action with a smaller absolute value but the same sign is also economic at the same inventory level. In other words, if a WS (respectively, BI) action is economic at a given inventory level, then a smaller (in absolute value) WS (respectively, BI) action is also economic at the same inventory level. Figure 16.2 illustrates these statements in stage 1: Specifically, the stated monotonicity in the WS case can be seen by comparing the functions $OC_1(a, 1)$ and $p_1(a)$; in the BI case, the stated monotonicity can be seen by comparing the functions $OC_1(a, 0)$ and $p_1(a)$.

Step 2: Mutually exclusive proper economic actions for a given inventory level. Consider a given stage. Given what argued in step 1, it should be clear that if at a given inventory level there exists a feasible proper economic BI (respectively, WS) action, then no feasible proper WS (respectively, BI) action is economic at this inventory level. In the BI case, compare the functions $OC_1(a, 1/4)$ and $p_1(a)$ in Figure 16.2. Moreover, if only the DN action is economic at a given inventory level, then no proper BI action and no proper WS action are economic at this inventory level. Compare the functions $OC_1(a, 1/3)$ and $p_1(a)$ in Figure 16.2. Hence, at each given inventory level, its corresponding feasible action set is partitioned into a connected subset of economic actions all with the same signs (possibly with the exception of the DN action), and a subset of noneconomic actions.

Step 3: Subdivision of the inventory set into economic BI/DN/WS subsets. It is now of interest to characterize the behavior of a given economic action when inventory increases in a given stage. To this aim, notice that in each stage the opportunity cost of a given feasible proper WS action (weakly) decreases when inventory increases, while the opportunity cost of a given feasible proper BI action (weakly) increases when inventory increases (assuming that the stated BI action remains feasible when inventory increases). In other words, in each stage the opportunity cost of a given feasible action rotates counterclockwise around the origin when inventory increases (in the BI case this is true so long as such an action remains feasible when inventory increases, as the domain where the opportunity cost is defined may also change when inventory increases). This should be interpreted in the weak sense, that is, the opportunity cost function may remain constant. In Figure 16.2, for proper WS action $a = -0.2$ compare functions $OC_1(-0.2, 1/4) = OC_1(-0.2, 1/3)$ and $OC_1(-0.2, 1)$; for proper BI action $a = 0.4$ compare functions $OC_1(0.4, 0)$, $OC_1(0.4, 1/4)$, and $OC_1(0.4, 1/3)$.

Because the immediate payoff function does not depend on inventory, it follows that in each stage if a feasible WS (respectively, BI) action is economic at a given inventory level, then only WS (respectively, BI) actions are economic in this stage at all higher (respectively, lower) inventory levels. Similarly, if only the DN action is economic at two different inventory levels in a given stage, this is also true in this stage at all the inventory levels in between them. But since at every inventory level only one type of action is economic, as argued in step 2, in each stage there exist nonoverlapping inventory subsets that share the same type of economic action and are separated by no more than two inventory levels $\underline{b}_j \leq \overline{b}_j$. Specifically, only BI actions are economic for $x \in [0, \underline{b}_j]$, only the DN action is economic for $x \in [\underline{b}_j, \overline{b}_j]$, and only WS actions are economic for $x \in [\overline{b}_j, 1]$.

Step 4: Subdivision of the inventory set into optimal BI/DN/WS subsets. The last step is to show that the inventory levels that subdivide the feasible inventory set into economic subsets are the optimal basestock targets. This is the only step that requires nongeometric reasoning. Consider stage j. Denote by $a_j^*(x)$ an optimal action in this stage at feasible inventory level x.

DN is the optimal action for each inventory level in set $[\underline{b}_j, \overline{b}_j]$, because DN is the only economic action for every inventory level in this set. Thus, it holds that $a_j^*(x) = 0, \forall x \in [\underline{b}_j, \overline{b}_j]$.

Consider an arbitrary inventory level x in the region where only BI actions are economic, that is, $x \in [0, \underline{b}_j]$. Finding an optimal action for this inventory level entails solving the following optimization problem:

$$\max_{a \in [0, \overline{C} \wedge (1-x)]} -s_j a + V_{j+1}(x + a)$$

This problem can be approached by ignoring the withdrawal and injection capacity limits, by finding an optimal solution $a_j^\diamond(0)$ to the capacity unconstrained problem:

$$\max_{a \in [0, 1-x]} -s_j a + V_{j+1}(x + a)$$

and by setting $a_j^*(0)$ equal to the minimum between $a_j^\diamond(0)$ and $\overline{C}$. The capacity unconstrained problem can be simplified by working with the ending inventory $y := x + a$, rather than an action a and the current inventory level x. That is, because $a \equiv y - x$, this problem is equivalent to:

$$\max_{y \in [x, 1]} -s_j y + V_{j+1}(y) + s_j x$$

Because, by assumption, at x only a BI action or the DN action can be optimal, the constraint $y \in [x, 1]$ can be relaxed to $y \in [0, 1]$ without loss of optimality. Thus, the resulting optimization is:

$$\max_{y \in [0, 1]} -s_j y + V_{j+1}(y) + s_j x$$

Further notice that the term $s_j x$ in the objective function of this problem does not affect the choice of an optimal solution, and it can be ignored in seeking such

TABLE 16.1 The Optimal Basestock Targets in Each Stage for the Illustrative Example

Basestock Target	Stage 1	2	3
BI	1/3	1	0
WS	1/3	1	0

a solution. Hence, the relevant optimization reduces to:

$$\max_{y\in[0,1]} -s_j y + V_{j+1}(y) \tag{16.4}$$

An optimal solution to this problem does not depend on which x in the set $[0, \underline{b}_j]$ one is considering. In particular, because DN is optimal for $x = \underline{b}_j$, it must be that $\underline{b}_j$ is an optimal solution to problem (16.4). Thus, it holds that the action $\overline{C} \wedge (\underline{b}_j - x)$ is optimal for every $x \in [0, \underline{b}_j]$, that is, $a_j^*(x) = \overline{C} \wedge (\underline{b}_j - x)$, $\forall x \in [0, \underline{b}_j]$. In other words, inventory level $\underline{b}_j$ is an optimal BI base-stock target.

It can be argued in a similar manner that inventory level $\overline{b}_j$ is an optimal WS base-stock target; that is, $a_j^*(x) = \underline{C} \vee (\overline{b}_j - x)$ for every inventory level x in set $[\overline{b}_j, 1]$. Thus, the optimality of a stage dependent double base-stock target structure has been established.

Table 16.1 displays the optimal base-stock targets in each stage for the illustrative example. The optimal policy is identical to that of the example discussed in Section 16.2 in the capacitated case.

16.5 Managerial Aspects

This section discusses some aspects of managerial relevance associated with the base-stock structure.

16.5.1 INVENTORY DEPENDENT PRICE CHARACTERIZATION

Suppose that in a given stage at least two types of actions are optimal at different inventory levels, that is, it does not hold that $\underline{b}_j = 0$ and $\overline{b}_j = 1$, in which case the DN action is optimal at all inventory levels, and that $\underline{b}_j = \overline{b}_j = 1$ or $\underline{b}_j = \overline{b}_j = 0$, in which cases filling up and emptying, respectively, the warehouse is optimal at all inventory levels. This implies that the stage j price does not have a unique interpretation: It can be interpreted as being *low*, *intermediate*, and *high*, respectively, at those inventory levels where a BI, DN, and WS action is optimal: only some of these characterizations apply if buying and injecting (respectively, withdrawing, and selling) is optimal at some inventory level but withdrawing and selling (respectively, buying, and injecting) is not optimal at any inventory level. This means that at a given decision epoch the characterization of a price must be made relative to the merchant's inventory level (Secomandi 2010).

16.5.2 INVENTORY INDEPENDENT PRICE CHARACTERIZATION

This discussion begs the question of whether there exist situations in which the characterization of a price in each stage is inventory independent, that is, if a BI, DN, or WS action is optimal at a given inventory level in a given stage, then the same type of action is optimal at all other inventory levels in the same stage. A situation when this occurs is the uncapacitated case, in which in every stage a warehouse can be filled up or emptied in a single stage (Charnes et al. 1966).

Interestingly, this is true even if the immediate payoff function is nonlinear in the action. In the capacitated case, the capacity functions that describe a feasible action as a function of inventory, that is, the quantity $\underline{C} \vee (-x)$ for a withdrawal and the quantity $\overline{C} \wedge (1 - x)$ for an injection, are nonlinear in the inventory level x. Instead, in the uncapacitated case these functions are linear, being, respectively, $-x$ and $1 - x$. Thus, the nonlinearity in the capacity functions in inventory has more profound implications in terms of the parameters of the optimal base-stock structure than the nonlinearity in the immediate payoff function in the action (Secomandi 2010).

16.5.3 OPTIMAL NONTRIVIAL CAPACITY UNDERUTILIZATION

In the example discussed in Section 16.4, as well in the capacitated version of that discussed in Section 16.2, in stage 1 it is optimal to underutilize the injection capacity at a nontrivial level. That is, it is optimal to utilize a positive amount of the injection capacity, but not its entirety, *whenever* BI is optimal, and an analogous statement is true *whenever* WS is optimal. Thus, the optimal merchant management of a commodity storage facility is far from trivial (Secomandi 2010).

16.5.4 COMPUTATION

The double base-stock target structure is very useful if one solves problems (16.1)–(16.2) by dynamic programming, because only two numbers need to be computed in each stage, that is, the optimal stage dependent base-stock targets. In this case, one still faces the issue that the state space is continuous. Hence, standard backward induction where one computes the value function for each feasible state in each stage is not easily applicable, because there is an infinite number of states in each stage. There are several possibilities to address this issue.

First, in addition to being concave, the optimal value function in each stage is piecewise linear. This means that it is characterized by an intercept and a finite number of slopes. Thus, one can recursively compute the intercept and the slopes of this function in every stage, from which one can then compute the optimal base-stock targets. Bannister and Kaye (1991) and Nascimento and Powell (2008) propose algorithms that exploit this structure (these authors deal with problems related to the warehouse problem).

Second, if the problem data satisfies a natural condition, then the points when the optimal value function may change slope, its breakpoints, can be determined *a priori*. Specifically, if the injection and the withdrawal capacities and the maximum inventory are all *integer* multiple of some real number Q, then the breakpoints are also so. Then, the only relevant inventory levels that need to be considered in each stage when solving problem (16.1)–(16.2) are 0, Q, $2Q$, ..., 1. Secomandi (2010) and Lai et al. (2010) exploit this result. This is also true in the example previously discussed, where $Q = 1/3$.

Third, one can solve the problem by linear programming. In this case, one does not obtain a policy for every stage and state, but only the optimal action for each state that is reached in every stage by following an optimal policy starting from a given initial state.

16.5.5 STOCHASTIC PRICE DYNAMICS

So far, this work has dealt with the case of deterministic price dynamics. In practice, commodity prices are uncertain. For example, commodity forward curves change in an uncertain fashion over time (the forward curve is the set of futures prices corresponding to different maturities). The problem considered here is relevant in practice because it is common among commodity storage managers to sequentially reoptimize a deterministic model similar to the one considered in this work to account for these changes in commodity forward curves (see, e.g., Maragos 2002). That is, in stage 1 one obtains an optimal policy given the current forward curve and implements the action corresponding to the current state in stage 1. After this action is implemented, one "moves" (in the real world) to the corresponding state at the beginning of stage 2. At this time, one also has a new forward curve available. One can then compute a new optimal policy corresponding to this new forward curve—this is the reoptimization step—and repeat the same procedure.

What this means is that one effectively computes a "slice" of a stage and forward curve dependent double base-stock target policy, that is, one in which the base-stock targets depend both on the stage and the forward curve in that stage. Because this is done by reacting to the evolution of the forward curve, in every stage one computes the base-stock targets only for those states that are visited by following such a policy; this is why one only computes a "slice" of this policy as opposed to the entire policy.

In the context of natural gas storage, Lai et al. (2010) show that the reoptimization version of the optimal deterministic policy described here is near optimal. In the same context, Secomandi et al. (2010) use Monte Carlo simulation to investigate how the choice of model for the evolution of the forward curve of this commodity affects the value of this reoptimized policy (see Seppi (2002) for a survey of commodity price evolution models). Moreover, the optimal policy when one models the stochastic evolution of the forward curve has a stage and forward curve dependent base-stock target structure, which generalizes the structure of the optimal deterministic policy discussed here (Secomandi et al. 2010).

16.5.6 FINANCIAL HEDGING

The uncertainty in commodity prices makes the merchant management of commodity storage a risky endeavor. Financial hedging of the price risk associated with the storage physical trading cash flows is common practice among those merchants for which this risk is costly. Delta hedging (Hull 2010, Chapter 6) is a basic risk management approach used by these merchants. Secomandi et al. (2010) show that it is possible to embed the reoptimized version of the optimal deterministic policy discussed here within a Monte Carlo simulation of the forward curve of a commodity to estimate the delta positions associated with this policy.

16.6 Conclusions

This work considers the warehouse problem, a foundational problem in the merchant management of commodity storage. The structure of the problem's optimal inventory trading policy is known to be of the base-stock target type. Motivated by the observation that the existing studies of this structure rely on marginal analyses that may not be easily accessible to managers, this work takes a different approach to establish the optimality of this structure. This allows one to proceed almost exclusively by relying on intuitive geometric arguments based on the notion of opportunity cost. Because the concept of opportunity cost is well known to merchants, if only qualitatively, it is hoped that managers will find this work of interest. Specifically, knowledge of this structure is important to them because it could be used to inform their inventory trading decisions. Alternatively, if they are using dynamic or linear programming models to support their inventory trading decisions, their decisions are likely consistent with the base-stock structure. In this case, this work may provide merchants with an enhanced understanding of the tools they use to support their decision making process.

REFERENCES

Bannister C.H. Kaye R.J. 1991. A rapid method for optimization of linear systems with storage. *Operations Research*, 39:220–232.

Bellman R. 1956. On the theory of dynamic programming—a warehousing problem. *Management Science*, 2:272–275.

Cahn A.S. 1948. The warehouse problem. *Bulletin of the American Mathematical Society*, 54:1073.

Charnes A., Drèze J., Miller M. 1966. Decision and horizon rules for stochastic planning problems: A linear example. *Econometrica*, 34:307–330.

Dreyfus S.E. 1957. An analytic solution of the warehouse problem. *Management Science*, 4:99–104.

Geman H. 2005. *Commodities and Commodity Derivatives: Modeling and Pricing for Agriculturals, Metals and Energy*. Chichester, UK:Wiley.

Hull J.C. 2010. *Risk Management and Financial Institutions*. Boston, Massachusetts: Prentice-Hall, second edition.

Lai G., Margot F., Secomandi N. 2010. An approximate dynamic programming approach to benchmark practice-based heuristics for natural gas storage valuation. *Operations Research*, 58:564–582.

Luenberger D.G. 1998. *Investment Science*. New York: Oxford University Press.

Maragos S. 2002. Valuation of the operational flexibility of natural gas storage reservoirs. In Ronn E., editor. *Real Options and Energy Management Using Options Methodology to Enhance Capital Budgeting Decisions*. London: Risk Publications, pages 431–456.

Massé P. 1946. *Les Réserves et la Régulation de L'Avenir Dans la Vie Économique*. Paris: Hermann.

Nascimento J.M. Powell W.B. 2008. Optimal approximate dynamic programming algorithms for a general class of storage problems. Princeton, New Jersey: Princeton University, Department of Operations Research and Financial Engineering.

Rempala R. 1994. Optimal strategy in a trading problem with stochastic prices. In: *System Modelling and Optimization*, volume 197 of *Lecture Notes in Control and Information Sciences*. Berlin: Springer, pages 560–566.

Secomandi N. 2010. Optimal commodity trading with a capacitated storage asset. *Management Science*, 56:449–467.

Secomandi N., Lai G., Margot F., Scheller-Wolf A., Seppi D. 2010. The effect of model error on the valuation and hedging of natural gas storage. Tepper Working Paper 2010-E71. Pittsburgh, Pennsylvania: Carnegie Mellon University.

Seppi D.J. 2002. Risk-neutral stochastic processes for commodity derivative pricing: An introduction and survey. In E. Ronn, editor, *Real Options and Energy Management Using Options Methodology to Enhance Capital Budgeting Decisions*, London: Risk Publications, pages 3–60.

Wallace S.W., Fleten S.-E. 2003. Stochastic programming models in energy. In A. Ruszczynski A., Shapiro A., editors, *Stochastic programming*, volume 10 of *Handbooks in Operations Research and Management Science*, Amsterdam, The Netherlands: Elsevier, pages 637–677.

Williams J.C. Wright B.D. 1991. *Storage and Commodity Markets*. Cambridge, UK: Cambridge University Press.

PART FIVE

Industrial Applications

CHAPTER SEVENTEEN

Procurement Risk Management in Beef Supply Chains

ONUR BOYABATLI, PAUL R. KLEINDORFER, AND STEPHEN R. KOONTZ

17.1 Introduction

The purpose of this chapter is to develop a basis for understanding the trade-offs facing a meat processing company (hereafter a "packer") in the choice of alternative arrangements for sourcing fed cattle, when that packer acts as a wholesaler into several final product markets. The general question posed is: what might influence a packer to source from long-term contracts versus spot markets as the basis for procurement of fed cattle when there are uncertainties and substitution possibilities in the demand for the resulting beef products supplied by the packer? Our focus is on the U.S. beef industry, which is the largest single industry within U.S. agriculture, generating between $34 and $37 billion per year in 2006–2008 and accounting for 20% of the annual total market value of agricultural products sold in the U.S. (USDA 2009). A similar analysis would apply to other cattle

Handbook of Integrated Risk Management in Global Supply Chains, First Edition. Edited by Panos Kouvelis, Lingxiu Dong, Onur Boyabatli, and Rong Li.

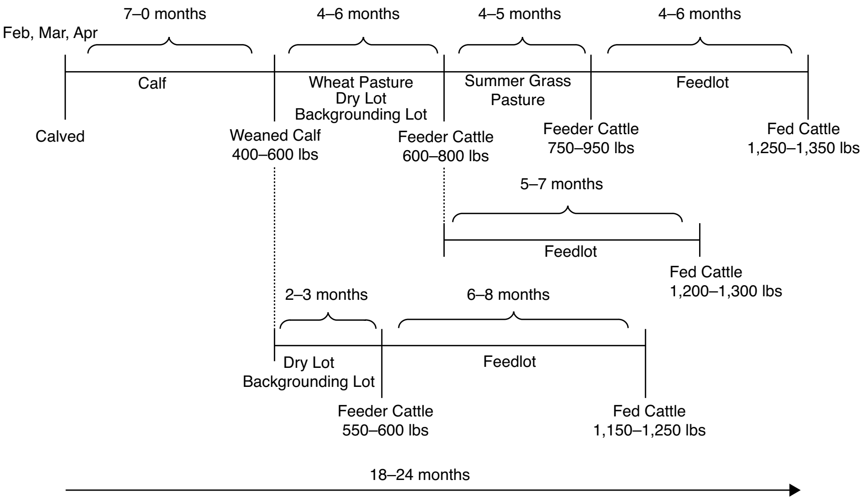


FIGURE 17.1 **Typical production system and timeline in beef supply chain** (GIPSA Report 2007).

producing regions of the world that rely for fed-cattle procurement on a mix of spot markets and long-term contracts (e.g., Europe and South America).

As shown in Figure 17.1, the beef industry is a combination of assembly and disassembly and of product flow smoothing. The base production unit in the industry—the beef cow herd—lives outdoors and consumes grass-based forage. The capital requirement in land is enormous and is the main reason why the cattle industry has not and will likely never integrate or consolidate. Beef cows produce a single calf per year and the large majority of calves are born in spring. Calves grow with the mother cow on grass pasture and are weaned in the fall. At this time the first major assembly occurs. Weaned calves are marketed through a multitude of auction barns and direct trade. Groups of calves are comingled and moved to so-called "backgrounding" operations. The purpose of backgrounding operations is to provide inexpensive animal growth on forage-based systems. Backgrounding operations include pasturing on growing winter wheat in the southern high plains, pasturing on stockpiled standing grasses, and feeding on inexpensive forages in confined operations. The length of backgrounding is highly variable, depending on the feeding regime. This variation in length of backgrounding is the primary means of smoothing the flow of cattle to packers.

The second major assembly occurs after backgrounding. After obtaining cheap growth of the animal frame, the animals are referred to as "feeder cattle" and are assembled by the cattle feeding industry. Feeder animals feed for 4–6 months depending on seasonal factors (such as energy requirements due to living outdoors and seasonal demand for beef consumption) and grain prices relative to beef prices. Finished animals are referred to as "fed cattle" and are marketed to packers.

As reported in the GIPSA Report (2007), there are some 25 large commercial fed cattle slaughtering and processing facilities in the U.S., and it is here that disassembly begins. Each animal can be used to produce a subset of hundreds of standard beef cuts. Further, excess fat is blended with lean beef trimmings—largely from the slaughter of beef animals, which include cull beef cows and bulls, to produce a number of beef products. These are packaged as premium products (program boxed beef) or commodity products (commodity boxed beef). Finally, each animal is used to produce a subset of by-products. The largest by-product is the hide, which is tanned for use as leather. The disassembly process continues through the beef distribution system. Food service such as restaurant chains may procure program beef. Grocery stores market a variety of commodity beef. There are distinct differences in regional demands across the U.S. and there is also a distinct variation in seasonal demands for types of beef products.

Beef markets have several interlinked markets that operate to determine pricing and delivery quantities at various stages along the supply chain. We will focus on the two markets of greatest interest to packers (see Figure 17.2):

1. The market between processors/packers and all upstream elements (including feedlots and prior elements) of the beef value chain;
2. The market between processors/packers and all downstream elements (including wholesalers and retailers) of the beef value chain.

Considering the upstream elements in the beef supply chain, there are actually two markets of interest: the spot market and the contract market.

Spot markets (also referred to as cash markets) are real-time regional markets for transactions of fed cattle, often through auctions. In keeping with the extensive literature on the subject (e.g., GIPSA Report 2007), we will assume throughout that spot markets are competitive, that is, the price is not sensitive to the actions of any of the agents (buyers or sellers) who participate in this market.

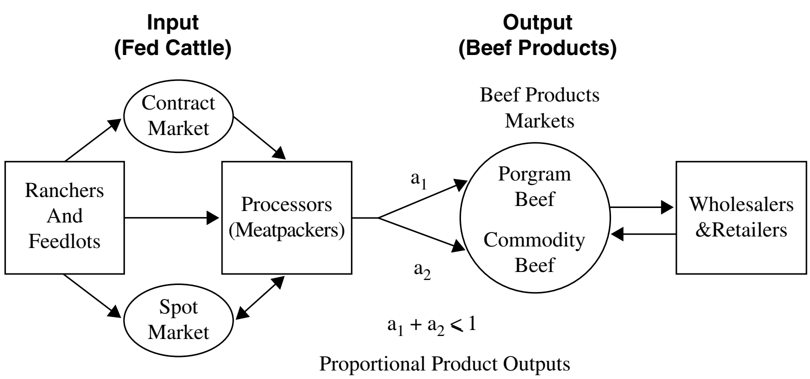


FIGURE 17.2 **Upstream and downstream elements for meatpackers in beef supply chain.**

Contract markets feature longer-term arrangements between feedlot owners and packers. The contracts themselves are often referred to as "marketing agreements." Such agreements may allow some flexibility in the quantity delivered, in the usual options form, or have more advanced features in pricing of yield risks (grid or formula-based) than fixed forwards based on (e.g., simpler live-weight metrics). The particular contract form analyzed below is the most common in the industry. It specifies the price per unit on the basis of the spot price prevailing at a specified market on delivery day. The usual form of this arrangement is that contract price equals spot price plus a fixed surcharge. The fixed surcharge is intended to cover the cost of additional feeding specifications that are part of the contract and, which give rise to the additional value of contract cattle resulting from the higher percentage of premium product (program beef) in these cattle. Contract cattle can also be resold in the spot market by the contracting packer, if they are not needed for production.

As noted in Kleindorfer and Wu (2003), in many organized commodity markets, a substantial portion of a seller's output or buyer's input is typically contracted for well in advance of delivery, with contract-based input in excess of 80% of total input, and where the spot market acts primarily as a topping up and hedge market. In contrast, for meatpackers in the U.S., the spot market is a very important source of physical supply, averaging for many meatpackers in excess of 60% of total supply according to GIPSA Report (2007), undertaken for the Grain, Inspection, Packers and Stockyard Administration (GIPSA) of the U.S. Department of Agriculture. This report provides the basic background and data for the computational experiments reported in this chapter. The heavy reliance on the spot market noted in the GIPSA Report is driven in part by the large number of small producers of cattle, who raise cattle as complements to their other farming activities, and the fact that spot sales in organized markets are an efficient way of bringing such cattle to market. Contract purchases obtained from larger feedlots offer certain advantages to packers such as the ability to contract for and monitor special feeding regimes that are intended to increase the quality of meat produced.

For the upstream market between a given packer and its suppliers (see Figure 17.2), the appropriate model would be one in which, following Wu and Kleindorfer (2005), a uniform product is provided by multiple suppliers characterized by heterogeneous costs (with quality differences captured in these costs as adjustments to the "full price" of a standard product). As our focus is on the integration of upstream and downstream markets, we will treat the upstream contract market as a single aggregate supplier, ignoring the details of how equilibrium price in this contract market is determined. We also assume that neither the cattle nor the finished products can be inventoried—they have a certain "ripe" or sale date towards which all contracting is directed.

For the downstream market between packers and buyers (see Figure 17.2), model features that are important include a multiproduct model (each unit of upstream product yields a certain number of units of saleable downstream products) with some quality differences between contract and spot purchases. Plant utilization is a critical issue for packers as their production technology (and our model) exhibits strong economies of scale.

Focusing on a single packer, we consider the optimal mix of contract and spot purchases in providing input from upstream feedlots and spot markets. Once delivered or purchased, these cattle are processed immediately and converted into the two beef products of interest, a premium product (program beef) and a standard product (commodity beef). As in the co-production literature (e.g., Bitran and Gilbert 1994), there is downward substitution in production in that all meat suitable for sale as premium product can be converted into the standard product. The downstream market into which the packer sells is price sensitive, and price is assumed to adjust to the quantity of both products sold into this market. As described in the GIPSA Report (2007), the market for beef products is competitive, so that the firm-specific price elasticity of demand for any given packer is large. At the market level, price adjusts quickly to clear all meat product and fed-cattle input markets. Further, meat product markets are closely related given the evident substitution effects between meat products. Imbalances in any individual market have impacts on other markets. Considerable volatility (both seasonal and product-based) exists in beef product and cattle markets and price-based clearance at the market level is critical.

In our companion paper, Boyabatlı et al. (2011), we develop the theoretical model and provide the optimal solution for the procurement portfolio of the packer.[1] This chapter describes the computational results for the above model based on data for the U.S. beef industry described in the GIPSA Report (2007), and complemented by industry demand and supply studies. Our analysis is focused on determining the impact on the optimal procurement portfolio of spot price and demand uncertainty, the degree of substitution between products in final markets, as well as the cost characteristics of the packer and the nature of quality and cost differences in the contract and spot markets.[2] As the focus is on the short and medium term, capacity and processing technology are assumed fixed.

This chapter intends to make the following contributions. It will provide insights about integrated risk management of input and output risks for the central player in the beef supply chains, the packer. Using a calibration based on the GIPSA Report (2007), the chapter will provide a foundation for understanding the complementary roles of contract and spot markets. In particular, the chapter will elucidate for the first time the value of contracting in the beef supply chain. As reviewed in the next section, this has been a point of considerable controversy in the policy debate concerning the structure and operations of the beef industry. In characterizing the structure of the optimal sourcing portfolio from a supply chain perspective, this chapter will provide an important contribution to the

[1] The theoretical model developed in Boyabatlı et al. (2011) focuses on a more general contract form, a special case of which is the marketing agreement contract analyzed in this paper.

[2] A part of these computational results are also reported in Boyabatlı et al. (2011) some of which are further generalized with analytical proofs.

ongoing debate on this issue. Beyond these contributions specific to the beef supply chain, our results will also indicate the value of integrated risk management across marketing, sourcing, and supply chain decisions.

This chapter proceeds as follows: We review relevant literature in the next section. Thereafter follows our model description in Section 17.3. Section 17.4 provides numerical simulations to illustrate the comparative statics of model results for processing, product market, and spot market parameters of interest. We conclude in Section 17.5 with a discussion of our managerial insights and the path forward for future research.

17.2 Literature Review

There is a rich literature in agricultural economics and operations management fields that considers supply chain contracting in the presence of spot markets. We refer the readers to Boyabatlı et al. (2011) for a review of the related literature from the operations management field. In this section, we will focus on the literature in the field of agricultural economics and management covering the beef industry. However, very little of this literature addresses supply chain management questions in a direct manner. This literature review will discuss some of the broader agricultural economic research, linking this to supply chain management questions addressed in this chapter. There are three broad areas of relevant literature: demand analysis, supply modeling, and the efficiency of pricing methods for marketing agreements.

Concerning the demand side of the beef markets, estimation of demand[3] elasticities are critical for market and policy analysis. Demand is inelastic so small changes in quantities result in relatively large impact at the market level. There is considerable volatility in livestock and meat prices. Further, meat demand is intrinsically variable. Red meat demand expanded considerably with the expanding U.S. economy and incomes during the 1960s and 1970s. However, health concerns and a number of other factors contributed to sharp declines in red meat demand following 1980. This decline in demand continued until 1998 and placed considerable economic pressure on the red meats industries.[4] Improving red meat demand in the late 1990s has been well documented (e.g., Marsh 2003). However, solid identification of the causes is not. The consumption of food away from home—or food prepared away from home—increases across the past years. Health-related concerns appear to be less, specialized preparation appears to be better, improvements in meat processing and technology appear to be better, or some combination, and have resulted in increased red meat demand along

[3] The industry standard for demand modeling is the Almost Ideal Demand System of Deaton and Muellbauer (1980). It is used in almost all the work referenced and has been found to produce elasticities with desirable forecasting properties (Kastens and Brester, 1996).

[4] See, for example, Braschler (1983), Chavas (1983), Dahlgran (1987), Moschini and Meilke (1989), Verbeke and Ward (2001) and Boetel and Liu (2003).

with the increase in food not-prepared at home. The GIPSA Report (2007) suggests that marketing agreement transaction methods (the contract market in our model) have emerged to address meat quality problems that are not addressable through the federal government developed grades and standards. The findings of these studies above are used to synthesize reasonable elasticities for program versus commodity beef in the numerical simulations reported in this chapter.

Concerning the supply side of beef markets, estimation of supply elasticities and the associated dynamic properties are critical for market and policy analysis. There are a large number of independent decision makers, the production process—the growth and development of beef animals—is lengthy, and the behavior by decision makers is in part anticipatory. A significant amount of literature has examined the dynamic properties of supply functions at the various stages of cattle and beef production.[5]

Another important area of the supply-related literature addresses technical progress within the beef industry and increased productivity. For example, the additional pounds of beef produced per animal in the breeding herd have increased 25% over the past 20 years. There are also been large changes in meat processing technology, changes and reductions in organized labor, and changes in provision of marketing service. These effects are slower but have substantial impacts on markets over time. The increase in productivity has maintained the total volume of beef production with a significant reduction in the size of the breeding herd.[6]

A final extensive and important supply-related literature addresses long-term investment in the cattle industry and the resulting cattle cycle dynamics (e.g., Schmitz 1997). There are inherent difficulties in modeling farm-level supply decisions and it may be that examining the herd building and liquidation decision itself is more useful. The cycle persists because there are cycle reinforcing actions and because expectations are still to a large part adaptive. The reinforcing actions are that when prices are relatively high and economic returns are favorable then returning additional young female animals to the herd and keeping additional cows in the herd exacerbates the high prices. Likewise, when prices are relatively low and economic returns are poor then selling young female animals into the meat production system and culling cows exacerbates the low prices. Further, expectation formation by beef producers have been found to be largely adaptive and not forward-looking (Antonovitz and Green 1990). Generally, the study of the cattle cycle is important but has provided no simple rules as far as the predictability of the cycle.

On the efficiency of pricing methods in the beef markets, the pricing mechanisms for alternative marketing arrangements[7] such as marketing agreements have been a more resent research interest in the agricultural economics literature. All of this research is focused on determining welfare implications to suppliers based

[5] Initial supply modeling work includes Reutlinger (1966) and Nelson and Spreen (1974), and the later work by Marsh (1983, 1984, and 1994).

[6] See, for example, Kuchler and McClelland (1989), Mullen et al. (1988), and Brester and Marsh (2001).

[7] The "alternative" refers to an alternative to the cash or spot market.

on the prospect of the exercise of market power by downstream procuring buyers. Comprehensive supply chain management issues and the optimal contracting behavior of the meatpacker—the focus of this chapter—have not been examined in detail in this literature.

Within producer groups, policy making and some government agency circles non-spot market procurement arrangements are referred to as captive supplies. These captive supplies are also referred to as contract supplies or marketing agreement cattle. These contracts are often more than simple forward contracts. Forward contracts comprise 5% of fed cattle transacted whereas the largest non-cash market arrangements are marketing agreements with formula pricing, in which the price paid for cattle is determined based on the amounts of each of type of beef actually present in the processed carcass. Large portions of participants within the beef industry have viewed such alternative marketing arrangements with skepticism and have often pushed for legislation to prohibit these arrangements. The most notable piece of legislation was the proposed Johnson Amendment to the 2000 Farm Bill. The amendment was not in the final bill, and a similar amendment was introduced but was not included in the 2008 Farm Bill, but there is persistent pressure by populist groups to limit or prohibit non-cash market transactions in the cattle industry.[8] We show in our chapter that, from the meatpackers perspective, this pressure is misplaced in that alternative marketing agreement (i.e., contract) that cattle are generally part of an efficient portfolio.

On the issue of competitive spot markets, Crespi and Sexton (2004) and Schroeder and Azzam (2003, 2004) provide a detailed examination of a classic dataset collected by the USDA Grain Inspection and Packers and Stockyards Administration. These data were comprehensive information collected in the Texas Panhandle area where captive supplies are a substantial proportion of total volumes and where some of the political pressure is the greatest. Market power was found to be present, but its economic consequences are minor to negligible. Like early structure-conduct-performance research on industrial organization, difficulties in interpreting the empirical research has lead to theoretical studies of the problem. Azzam (1998) is one of the earliest studies and determines that the price impacts of captive supplies are ambiguous due to relative changes in supply and demand of spot market and non-spot market animals. Zhang and Sexton (2000, 2001) examine the role of transportation costs as a source of market power. Xia and Sexton (2004) examine a theoretical model of top-of-the-market contract pricing clauses that are most often used with alternative marketing agreements. Wang and Jaenicke (2006) is the most recent research supported by results derived through simulation. The authors find that impacts of captive supplies on cash market price are ambiguous. While all of these studies find the potential for

[8] There is a similar but weaker movement related to the use of non-cash market arrangements in the pork-hog industry where the volume of these arrangements is more than double that in the beef industry—based on the proportion of total industry volume accounted for by non-cash arrangements. Policy makers appear to treat the issue within the beef industry as unique to the beef industry and do not recognize that the practice of reliance on both contract and spot markets is persistent in almost all commodity industries, as discussed in Kleindorfer (2008).

market power through the strategic use of non-cash market arrangements, few examine the potential efficiency benefits that may come with improved supply chain management (the focus of this chapter). The exception is Love and Burton (1999) who build a model of captive supplies where the packing firm has declining average costs of processing with its processing facilities and an incentive to backward integrate to assure adequate supply to take advantage of its economies of scale.

Against the background of the above literature, we can note several important lacunae. For the upstream market, there is no research on the optimal mix of procurement methods (contract vs. spot) within the beef industry. Furthermore, the key issue of quality/yield risks (which are different across contract and spot procurement methods) needs to be addressed and integrated with production and demand management. For the downstream market, the key issue that needs to be addressed is that of multiple products arising from processed beef (premium and standard products) and the demand uncertainties and substitution effects associated with these. It is precisely on these key issues, and their related impacts on optimal processing decisions for the producer (here the meatpacker), that we focus our model and our results.

17.3 Model Description

This section describes our modeling framework that is developed in Boyabatlı et al. (2011). We consider a packer that procures and processes fed cattle to produce two beef-products, a premium (program beef) and a standard (commodity beef) product. We model the packer's procurement, processing and production decisions in a two-period framework. Before discussing the details of these decisions, we provide some notations that we will use throughout the chapter. A realization of the random variable $\tilde{y}$ is denoted by y. $\mathbb{E}$ denotes the expectation operator, and bold face letters represent vectors of the required size. Vectors are column vectors and $'$ denotes the transpose operator. Monotonic relations are used in the weak sense unless otherwise stated. We use "C-cattle" to denote the cattle sourced from the contract market and "S-cattle" to denote the cattle sourced from the spot market.

17.3.1 PROCUREMENT DECISION

In line with the above discussion, we consider two sources for procurement, marketing agreement contracts and spot markets. The marketing agreement contract specifies the number of C-cattle that are committed by the packer in advance of the spot market and are delivered to the packer on the spot day. The packer can also buy S-cattle from the spot market on the day. Let Q^C denote the number of C-cattle and $Q^S(P^S)$ denote the number of S-cattle at the prevailing spot price P^S. We assume that $\tilde{P}^S$ follows a normal distribution with mean μ_S and standard deviation σ_S.

There are differences between C- and S-cattle in terms of meat quality, processing cost and contract price. Processing C-cattle is cheaper and leads to a higher yield of carcass meat suitable for producing the premium product (where the additional yield is denoted as Δ). We will discuss these differences in detail later in this section. C-cattle are priced through formula (i.e., grid) pricing that tie the base price to publicly reported spot prices and specify surcharge for high quality meat (MacDonald 2003). In line with this, in our model, the unit price of C-cattle is $P^S + v\Delta$ and is based on the prevailing spot price at the time of the delivery plus a surcharge ($v\Delta$) to reflect the higher quality of C-cattle. The unit price of S-cattle is the prevailing spot price P^S with an additive transaction cost $t > 0$ applied. This transaction cost reflects transportation cost from the auction barn (spot market) to the packers plant and weight loss between purchase and processing. The packer can also sell C-cattle that it receives under contract in the spot market. The unit sales price is $(1 - \omega)P^S$ where $0 < \omega < 1$ represents a transaction cost.

17.3.2 PROCESSING DECISION

Fed-cattle processing has two main characteristic features. First, packers have high incentives to increase plant utilization due to significant scale economies (Ward and Schroeder 2002). Second, animal nonuniformity creates frictions in cattle processing (Hennessy 2005); and C-cattle are more uniform than S-cattle (Hayenga et al. 2000). We define $\mathbf{z}' = (z^C, z^S)$ as the vector of processed cattle composed of C-cattle, z^C, and S-cattle, z^S. We assume that there exists a physical processing capacity constraint K (hereafter referred as plant size) such that $\mathbf{1}'\mathbf{z} \leq K$; and the total processing cost is denoted by $C(\mathbf{z}) = c_0\mathbf{1}'\mathbf{z} + \delta z^S + c_1(K - \mathbf{1}'\mathbf{z})^2$. Here, c_0 is the common processing cost parameter, $\delta > 0$ represents the additional processing cost of S-cattle due to animal nonuniformity and c_1 is a (quadratic) utilization cost parameter. As the total processed cattle ($\mathbf{1}'\mathbf{z}$) increases, the average unit cost $\frac{C(\mathbf{z})}{\mathbf{1}'\mathbf{z}}$ decreases. In addition to the volume-variable costs, fixed costs are also important elements of the packer cost structure. They represent payments to capital providers and indirect facility costs. We neglect these in the model development as they do not affect the optimal solution. Fixed costs are reflected in the calibration underlying our numerical results in Section 17.4. Decreasing short-term average costs throughout the entire range of feasible input levels are well documented and important for packers in the beef industry (Koontz and Lawrence 2010).

17.3.3 PRODUCTION DECISION

In the beef supply chain, beef products are grouped into two major categories, program beef and commodity beef. Program beef is the premium product. In our model, product 1 refers to program beef and product 2 refers to commodity beef. Each unit (head) of processed cattle leads to carcass capacities in fixed proportions that can be used for production.

We denote a_i^j as the fixed proportion of the carcass for product $i = \{1, 2\}$ from cattle type $j = \{C, S\}$. We assume $\mathbf{a_1'} = \left(a_1^C, a_1^S\right) < \mathbf{a_2'} = \left(a_2^C, a_2^S\right)$, that is, carcass capacity is lower for the premium product than for the commodity product, whatever the source of the carcass. We also assume $a_1^j + a_2^j = s \leq 1$ for $j \in \{C, S\}$, that is, the total carcass yield is identical for both cattle types and there could be yield losses in processing ($s < 1$). To capture the quality difference, we assume $a_1^C = a_1^S + \Delta$ and $a_2^C = a_2^S - \Delta$ where $\Delta \geq 0$ denotes the quality difference of C-cattle. C-cattle have a higher carcass capacity for the premium product. Since the total carcass capacity is fixed, the proportion of the standard product is higher with S-cattle.

The firm-specific demand for beef products is stochastic, price-dependent, and represented by the linear inverse-demand functions $p_1(\mathbf{x}, \xi_1) = \xi_1 - b_1 x_{11} - e(x_{22} + x_{12})$ and $p_2(\mathbf{x}, \xi_2) = \xi_2 - b_2(x_{22} + x_{12}) - e x_{11}$. Here, $\mathbf{x}' = (x_{11}, x_{22}, x_{12})$ is the production vector, e represents the cross-price elasticity parameter and ξ_i, b_i, p_i denote the market size, own price slope of the demand function, and price for product i respectively. In the production vector $\mathbf{x}$, x_{kl} denotes the quantity of product l produced from the meat capacity ($\mathbf{a_k'z} = a_k^C z^C + a_k^S z^S$) dedicated to product k. We assume that $\boldsymbol{\xi}' = (\xi_1, \xi_2)$ follows a bivariate normal distribution with mean vector $\bar{\boldsymbol{\xi}}' = (\mu_1, \mu_2)$ and covariance matrix Σ, where $\Sigma_{ii} = \sigma_\xi^2$ and $\Sigma_{ij} = \rho_\xi \sigma_\xi^2$ for $i \neq j$ and ρ_ξ denotes the correlation coefficient. Since the first product is premium product, we have $\mu_1 > \mu_2$ (i.e. for identical quantities), the expected price of the first product is higher; and $b_1 > b_2$ (i.e. the first product demand is less responsive to changes in price than the second product). In particular, we assume $b_2 < b_1 \frac{a_1^S}{a_2^S}$. This is an appropriate assumption for beef markets where price sensitivity is considerably higher for premium products than for standard products.

We allow for two different substitution channels for production. There exists *demand substitution* through the cross-price elasticity parameter e. Since beef-products are natural substitutes, the price of each product is decreasing in the price of the other product ($e > 0$) and this cross-price effect is lower than the own-price effect ($e < \min(b_1, b_2)$). There is also *downward product substitution*: the packer can produce standard product using the carcass capacity dedicated to premium product, and not vice versa. We assume that the packer uses a market clearing pricing strategy, that is, all the available carcass is processed into one of other of the two beef products and price is adjusted in profit-maximizing fashion to sell all finished products.

17.3.4 THE MODEL

We model the packer's decision problem as a two-stage stochastic recourse problem. In stage 0, the packer decides on the number of C-cattle (Q^C) to contract with respect to spot price $\tilde{P}^S$ and product market $\tilde{\boldsymbol{\xi}}$ uncertainties. At stage 1, these uncertainties are realized and Q^C is delivered to the packer. The packer decides on the number of cattle to buy from the spot market (Q^S), the number of cattle to process out of the available S-cattle (z^S) and C-cattle (z^C), the number of cattle to sell back to the spot market ($Q^C + Q^S - z^C - z^S$) and the production quantities of two beef products that either come from their dedicated carcass capacities (x_{11}, x_{22}), or through substitution of the premium product carcass capacity to produce standard product (x_{12}). The objective of the packer is to maximize the expected total profit at stage 0.

We now formulate the packer's decision problem starting from stage 1:

$$\begin{aligned}
\max_{Q^S,\mathbf{z},\mathbf{x}} \; & -Q^C(P^S + v\Delta) - Q^S(P^S + t) + (1-\omega)\,P^S\left[Q^C + Q^S - \mathbf{1}'\mathbf{z}\right] \qquad (17.1)\\
& - \left[c_0\mathbf{1}'\mathbf{z} + \delta z^S + c_1(K - \mathbf{1}'\mathbf{z})^2\right]\\
& + x_{11}\,(\xi_1 - b_1 x_{11}) + (x_{22} + x_{12})\,(\xi_2 - b_2\,(x_{22} + x_{12})) - 2e\,(x_{22} + x_{12})\,x_{11}\\
\text{s.t. } & z^C \le Q^C, \quad z^S \le Q^S, \quad \mathbf{1}'\mathbf{z} \le K\\
& x_{11} + x_{12} = \mathbf{a_1'z}, \quad x_{22} = \mathbf{a_2'z}\\
& Q^S \ge 0, \quad \mathbf{z} \ge \mathbf{0}, \quad \mathbf{x} \ge \mathbf{0}
\end{aligned}$$

In (17.1), the first two terms represent the total procurement cost of the packer. The third term is the revenue from spot market sales and the fourth term is the total processing cost of the packer. The final terms in the objective function denote the sales revenue from the beef products. The first two constraints ensure that the packer does not process more than the available capacity of a particular cattle type. The third constraint guarantees that the packer processes within plant size. The fourth and the fifth constraints represent the available carcass capacity for each beef product under market clearing pricing strategy. Let $\Pi(Q^C; P^S, \boldsymbol{\xi})$ denote the optimal stage 1 profit for a given Q^C.

Anticipating these decisions, at stage 0, the packer solves for the optimal number of C-cattle to contract, Q^{C*}, to maximize the expected firm profit: $V^* = \max_{Q^C \ge 0} \mathbb{E}\left[\Pi(Q^C; \tilde{P}^S, \tilde{\boldsymbol{\xi}})\right]$ where the expectation is taken over $\tilde{P}^S$ and $\tilde{\boldsymbol{\xi}}$. We assume that the distributions of these two random variables are statistically independent. To deal with the non-negativity of the market price, we assume that the coefficient of variations are not extremely large, and hence, the effect of negative values is negligible.

We refer the reader to Boyabatlı et al. (2011) for the explicit characterization of the optimal contracting decision. We close this section with an important observation about the efficiency of contract market in the beef supply chain. As reported in Hayenga et al. (2000), packers note the following factors driving contract-market procurement: (1) risk of not being able to obtain cattle from the spot market, (2) nonuniformity of S-cattle and corresponding higher processing costs, and (3) higher quality of C-cattle over S-cattle. In parallel with this empirical observation, in our model, it is straightforward to show that if there is no spot procurement transaction cost ($t = 0$), no additional processing cost for S-cattle ($\delta = 0$) and no quality difference between C-cattle and S-cattle ($\Delta = 0$), then the packer does not contract any C-cattle. In the next section, we shed more light on the the main drivers of the optimal procurement portfolio as well as on several performance measures using numerical experiments.

17.4 Computational Experiments for the Beef Supply Chain

This section describes computational results for the above model. Our primary objective is to provide insights on some fundamental intuitions about the optimal integration of upstream contracting and downstream demand management. This section is calibrated on the typical packer, in terms of size and cost characteristics, described in the GIPSA Report (2007), thus allowing further insights into some of the controversies surrounding that important study. The GIPSA data on packer characteristics were complemented by industry demand and supply studies. The GIPSA data pertain to the U.S. beef industry for the period October 2002 through March 2005. We focus on an average sized plant (see Tables 3.2, 3.3 and Figure 3.1 of the GIPSA Report) with rated capacity of 25,000 head of cattle per week (corresponding to the mean plant size of the GIPSA Report of 103,733 cattle per month as reported in Table 3.2). Tables 17.1, 17.2 and 17.3 provide the benchmark values for this packer and the relevant range for the sensitivity analysis.

The mean spot price μ_S is set to be \$1600 (per head) and is in line with the average auction barn (spot market) price of \$1.32 per pound (with an average weight of 1200 lbs per head) as reported in Table 5.1 of GIPSA Report. We set σ_S, spot price variability to 8% of μ_S and is consistent with the reported variability of average weekly prices in Table 5.1 of GIPSA Report. The surcharge paid for the quality difference of C-cattle, $v\Delta$, is set such that the average procurement price of C-cattle and S-cattle are identical as follows from Table 5.1 of GIPSA Report (the average price of C- and S-cattle are reported as \$1.32 per pound).

The GIPSA data on packer characteristics were complemented by industry demand and supply studies. For example, t, transaction cost in spot procurement, is set to be 4% of the mean spot price μ_S. This 4% represents the pencil shrink on the cattle purchased from the spot market. The shrink is the water loss in

TABLE 17.1 Description of the Spot and Contract Market Characteristics in Numerical Studies

Spot and Contract Market Characteristics			
Notation	**Description**	**Benchmark Value**	**Range**
ω	Transaction cost in spot sales (percentage)	4% of P^S	0% to 4% of $\bar{P}^S$ with 0.5% increments
t	Transaction cost in spot procurement	4% of μ_S (\$64/head)	
μ_S	Mean spot price	\$1600/head	
σ_S	Spot price volatility	8% of μ_S (128)	4% to 9% of μ_S with 1% increments
v	Surcharge parameter for quality difference of C-cattle	(\$4800/head) Δv=3.75% of μ_S	2.5% to 4.25% of μ_S for surcharge ($v\Delta$) with 0.25% increments

TABLE 17.2 Description of the Processing Characteristics in Numerical Studies

Processing Characteristics			
Notation	**Description**	**Benchmark Value**	**Range**
c_1	Utilization cost parameter	\$0.001	0, 0.001, 0.003, 0.005, 0.01, 0.015, 0.025, 0.05, 0.1
c_0	Common processing cost parameter	\$100/head	0 to 250 with 50 increments
δ	Nonuniformity cost of S-cattle	\$1.39/head	0 to 2.78 with 0.695 increments
K	Plant size	25000 head/week	

the animal between the feedlot and packing plant. Since C-cattle tend to be produced close to the plant, and the shrink is far less than the market driven 4% and is reasonably close to (and set to be) zero.

TABLE 17.3 Description of the Product Market Characteristics in Numerical Studies

Product Market Characteristics			
Notation	**Description**	**Benchmark Value**	**Range**
e	Cross-price elasticity parameter	0.005	0 to 0.01 with 0.0025 increments
b_1	Own price coefficient for program beef	0.035	
b_2	Own price coefficient for commodity beef	0.01	
μ_1	Mean demand of program beef	3800	
μ_2	Mean demand of commodity beef	3000	
$\sigma_{\xi_1} = \sigma_{\xi_2} = \sigma_\xi$	Demand variability	6% of μ_2 (180)	3% to 8% of μ_2 with 1% increments
ρ_ξ	Demand correlation	0.9	0.75 to 1 with 0.05 increments
a_1^S	Fixed proportion of program beef with S-cattle processing	0.18	
a_2^S	Fixed proportion of commodity beef with S-cattle processing	0.42	
Δ	Quality difference $= a_1^C - a_1^S = a_2^S - a_2^C$	0.0125	0 to 0.015 with 0.0025 increments
s	Total proportion of usable carcass $= a_1^C + a_2^C = a_1^S + a_2^S$	0.60	

On the cost calibration, we focus on 25,000 head cattle processing, 50% of which comes from the spot market (as consistent with the GIPSA Report). We calculated average total cost (ATC), that is total processing cost divided by the total quantity processed, at 95%, 75%, and 50% utilization rates. The benchmark ATC number (at 95% utilization) is \$139 and is taken from Table 3.1 of the GIPSA Report. The cost estimation in the GIPSA Report illustrates that a plant operating at 75% utilization rate has an ATC that is 6% higher than the benchmark ATC; and a plant operating at 50% utilization rate has an ATC that is 14% higher than the benchmark ATC. Moreover, the increase in ATC is more significant at lower utilization rates. The nonuniformity cost δ corresponds to the 1% of the benchmark ATC. Finally, fixed facility costs of $900K$ per week were assumed, representing fixed staffing and maintenance costs and payments to investors, which is representative of the range of fixed costs of medium-sized U.S. plants. To determine the final calibration, we minimized the sum of the quadratic difference between the estimated and the specified target ATC values at 95%, 75%, and 50% utilization rates. The resulting cost parameters (fixed cost, δ, c_0, and c_1) provide a good fit to the observed pattern above.

On the demand calibration, since the beef product demands are highly correlated, we set ρ_ξ to be 0.9. Since demand variability is lower than the spot price variability, as consistently observed in the beef markets, we set σ_ξ to be 6% of the mean demand of the standard product (μ_2). The demand parameters, own price coefficients b_1 and b_2, and cross-price elasticity parameter e, are set to be sufficiently low such that the firm-specific price elasticity of demand is large. With the resulting set of parameters, the expected beef price (calculated from expected price of each product rated by its corresponding fixed proportion) is calculated as \$2.60 per pound. This is consistent with the average beef price reported in Table 1.4 of the GIPSA Report (the reported gross price is \$2.62 and the net price is \$2.57). The expected profit of the packer is calculated as \$2.04 million per week, and corresponds to 5.6% of the total sales revenues from two beef markets. These two profit measures are representative features of a medium-size packer in the beef industry.

As final validation tests of the model calibration, we analyze the optimal sourcing portfolio, expected utilization and the expected spot selling of the C-cattle. As depicted in Figure 17.3, in the period of the GIPSA study (October 2002 to March 2005), the ratio of spot procurement is higher than, yet close to, the contract procurement. At the benchmark parameter values, the optimal sourcing portfolio is composed of 41.6% contract market procurement and 58.4% spot market procurement. This is consistent with the observed pattern in Figure 17.3.

The expected utilization of the packer is calculated to be 77% and the expected spot sales ratio (the ratio of expected spot number of C-cattle sold back to the spot market to the total C-cattle) is 2.2% (i.e. the packer almost always uses C-cattle for processing). These two numbers are also consistent with the characteristics of a medium-size packer in the beef industry.

For computational experiments, we programmed the first-order-condition and the other performance measures in MATLAB. We validated the code against a number of tests that included making comparisons between the MATLAB results

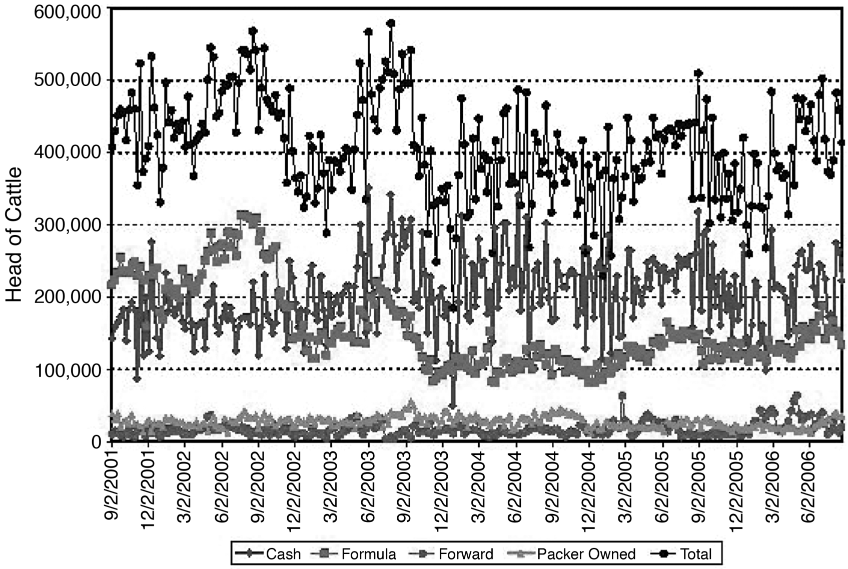


FIGURE 17.3 **Sourcing classification of fed cattle procurement in beef supply chains. Here, "cash" refers to spot market procurement and "formula" refers to the marketing agreement contracts.**

and (1) explicitly calculated optimal values for the performance measures when $\tilde{\xi}$ and $\tilde{P}^S$ equaled their mean values (in this case, σ_ξ and σ_S were assigned very low values so that all the probability mass was located at the mean); (2) results of several special cases of the problem for which we analytically know the behavior of the optimal performance measures (for example, for $\omega = v = 0$ we have $Q^{C^*} = K$), and (3) a number of comparative static results that can be proven analytically (e.g., Q^{C^*} is decreasing in ω).

A number of performance measures were computed for the experiments reported here, all of them evaluated at the optimal solution to the packer's expected profit maximization problem. Specifically, we report:

PERF-1. *The optimal volume of C-cattle to contract:* Q^{C^*}

PERF-2. *Expected spot procurement at the optimal solution:* $\mathbb{E}[Q^{S^*}]$

PERF-3. *Optimal portfolio (contract intensity) ratio:* $\frac{Q^{C^*}}{Q^{C^*}+\mathbb{E}[Q^{S^*}]}$

PERF-4. *Expected optimal profit of the packer:* $\mathbb{E}[\Pi^*]$: This includes \$900,000 in fixed costs (including payments to owners/investors) per week.

PERF-5. *Value of contract market:* $\frac{\mathbb{E}[\Pi^*(Q^{C^*})]-\mathbb{E}[\Pi^*(0)]}{\mathbb{E}[\Pi^*(Q^{C^*})]}$. This captures the relative value loss between the packer using the optimal number of contracts and the packer not using any contracts.

PERF-6. *Value of spot market:* $\frac{\mathbb{E}[\Pi^*(Q^{C^*})]-\mathbb{E}[\Pi^*(Q^{C^*}|t\to\infty,\omega\to 1)]}{\mathbb{E}[\Pi^*(Q^{C^*})]}$. This captures the relative value loss between the packer using the optimal number of contracts (with spot involvement) and the packer using the optimal number of contracts (without spot involvement).

PERF-7. *Expected capacity utilization of the packer's plant:* $\frac{\mathbb{E}[z^*]}{K}$ where z^* denotes the optimal processing volume at stage 1.

To illustrate the impact of the various parameters of interest from Tables 17.1 to 17.3, we first compute the elasticity of the performance measure with respect to each of the parameters, for a variation of ±5% around the benchmark case. Elasticity of performance metric "F" w.r.t. parameter "p" is defined as $\frac{\partial F}{\partial p} \times \frac{p}{F}$, and therefore represents the percentage change in F arising from a one percentage point change in p. The results of this exercise are shown in Table 17.4. Second, we numerically analyze the impact of these parameters over their entire range as specified in Tables 17.1 to 17.3. The arrows in the cells in Table 17.4 indicate these results. Some of these results are non-monotonic. In these cases, we demonstrate the impact with multiple arrows in the order of observation as the parameter of interest increases. For example, ↓↑ implies that the particular performance metric first decreases then increases with an increase in the parameter of interest.

TABLE 17.4 Impact of Parameters on the Performance Measures

	Contract PERF-1	Spot PERF-2	Portfolio PERF-3	Profit PERF-4	C- Value PERF-5	S- Value PERF-6	Utilization PERF-7
c_0	−1.13804 ↓	0.419687 ↑↓	−0.90230 ↓	−0.59950 ↓	−1.41721 ↓	2.11155 ↑	−0.25744 ↓
e	−0.19233 ↓	0.035710 ↑	−0.13331 ↓	−0.10353 ↓	−0.14660 ↓	0.48714 ↑	−0.05842 ↓
δ	0.10003 ↑	−0.062454 ↓	0.09516 ↑	−0.00507 ↓	0.26556 ↑	−0.02956 ↓	0.00336 ↑
Δ	−3.72820 ↓	3.086226 ↑	−3.52313 ↓	−0.13318 ↓	−6.58813 ↓	2.35207 ↑	−0.26155 ↑↓
ω	−0.48103 ↓	0.148112 ↑	−0.46067 ↓	−0.00347 ↓	−0.80789 ↓	−0.01457 ↓	−0.00913 ↑
v	−6.69295 ↓	3.439525 ↑	−6.56454 ↓	−0.08706 ↓	−8.83006 ↓	2.23974 ↓	−0.31326 ↑
c_1	0.42442 ↑	−0.194364 ↓	0.36397 ↑	−0.02799 ↓	0.97425 ↑	−0.16225 ↓↑	0.05766 ↑
σ_S	−2.07179 ↓	0.614115 ↑	−1.93687 ↓	0.19425 ↑	−3.86640 ↓	0.84328 ↑	−0.15958 ↓
σ_ξ	−0.99127 ↓	0.545250 ↑	−0.87836 ↓	0.13939 ↑	−2.04259 ↓	0.84293 ↑	−0.12723 ↓
ρ_ξ	−0.20042 ↓	0.103052 ↑	−0.17646 ↓	0.02742 ↑	−0.43698 ↓	0.16120 ↑	−0.02517 ↓

As an example, consider the impact of Δ on Q^{C*} in Table 17.4. The elasticity of Q^{C*} w.r.t. Δ is given as -3.72820. Noting the linear approximation being used here to estimate elasticities, this means that, in the neighborhood of the Base Case, an increase in Δ of 1% would lead to a 3.72820% decrease in Q^{C*}, *ceteris paribus*. This monotonic behavior is also observed over the entire range of Δ as depicted by $\downarrow$. An increase in Δ has two effects: first, it increases the fraction of premium product in C-cattle with positive profit impacts given the higher price for the premium product; second, it increases the surcharge paid over the spot price for C-cattle (with the surcharge equal to $v\Delta$). Given the value of v in the market, the second effect dominates the first in our numerical experiments.

As can be seen further in Table 17.4, an increase in Δ would lead to an increase in spot procurement, a decrease in the contract intensity ratio, a decrease in expected profits, a decrease in the value of the contract market, and an increase in the value of the spot market. The impact on the expected capacity utilization is non-monotonic. When Δ (thus, the surcharge) is sufficiently low, the firm contracts up to the plant size. In this case, expected processing quantity $\mathbb{E}[z^*]$ increases in Δ as higher price for the premium product induces the packer to process more of C-cattle (and sell less of it to the spot). Therefore, expected utilization increases. When Δ is sufficiently high, the firm does not contract up to plant size. In this case, a higher Δ decreases Q^{C*} and expected utilization decreases.

Rather than dwell on the rationale and intuition for each of the results shown in Table 17.4, we focus on the effects of input and output price variability, contract market transaction costs, quality difference between C- and S-cattle, utilization cost parameter, and product and demand substitution.

17.4.1 EFFECT OF SPOT PRICE AND PRODUCT MARKET VARIABILITY

In this section, we analyze the effect of spot price variability (σ_S) and product market variability (σ_ξ, ρ_ξ) on the key performance indicators. For brevity, on the impact of product market variability, we will only provide figures for σ_ξ.

As depicted in panel A of Figure 17.4, Q^{C*} decreases in σ_S. In our numerical experiments, we observe that the packer almost never sells C-cattle back to the spot market. Therefore, the impact of the spot price variability on the optimal contract volume is through its impact on the spot procurement. Since the packer only buys from spot market when spot price is sufficiently low, with a higher σ_S, the packer benefits from low spot price realizations by procuring S-cattle cheaper, whereas the packer is not affected from the high spot price realizations. Therefore, the packer's reliance on S-cattle increases, and in turn, Q^{C*} decreases.

For the effect of σ_ξ and ρ_ξ on Q^{C*}, we note here that a higher σ_ξ or ρ_ξ increases the variability of product market returns. For ρ_ξ, this is because a higher correlation decreases the diversification benefit from operating in two markets. Since C-cattle is always processed (and is not sold back to the spot market), the change in the variability of product market returns does not have an impact on the expected marginal value of processing the C-cattle. On the other hand, the packer

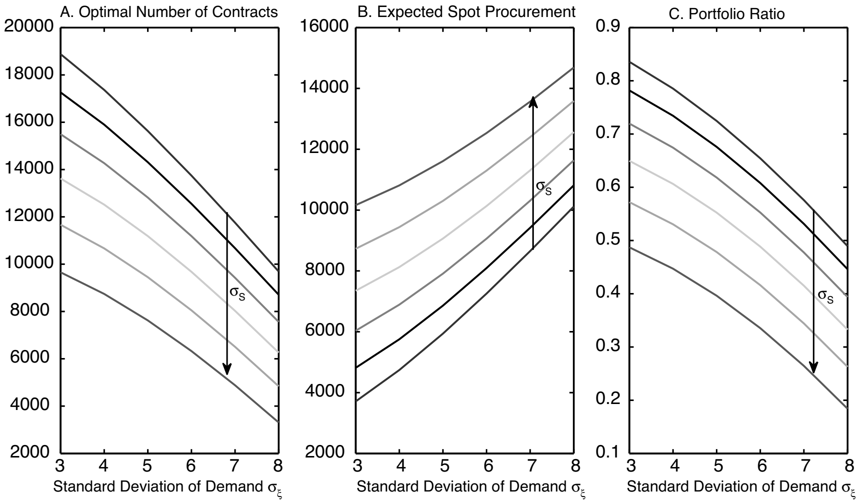


FIGURE 17.4 **Impact of spot price variability (σ_S) and product market variability (σ_ξ) on the optimal procurement portfolio: σ_S ranges from 4% to 9% of the mean spot price (μ_S) with 1% increments and σ_ξ ranges from 3% to 8% of the mean demand of the standard product (μ_2) with 1% increments.**

processes S-cattle (after all the C-cattle is processed) only if the product market return is sufficiently high. In other words, with a higher σ_ξ or ρ_ξ, the S-cattle processing benefits from the higher variability in product market returns. Since the packer relies more on the S-cattle, Q^{C*} decreases. The result with respect to σ_ξ is depicted in panel A of Figure 17.4.

We now analyze the effect of variability on the expected spot procurement. As depicted in panel B of Figure 17.4, with a higher σ_S or σ_ξ, expected spot procurement increases. The same holds true with an increase in ρ_ξ. These results are driven by two effects: First, S-cattle processing benefits from a higher σ_S (a higher σ_ξ or ρ_ξ). This is because the packer optimally processes S-cattle only when the spot price is sufficiently low (or the product market return is sufficiently high). Second, Q^{C*} decreases and the packer relies more on the spot procurement. As Q^{C*} decreases and the expected spot procurement increases, the optimal portfolio ratio decreases in σ_S and σ_ξ as depicted in panel C of Figure 17.4. The same holds true with an increase in ρ_ξ.

For the impact on the expected profit, we first analyze the effect of σ_S. The packer has two options on the spot market: spot selling and spot procurement. As we pointed out above, expected spot selling is very small in the optimal solution within our numerical setting. Since the packer optimally procures from the spot market only if the spot price is sufficiently low, the value of spot procurement increases in σ_S. Therefore, the expected firm profit increases in σ_S as depicted in panel A of Figure 17.5. The effect of σ_ξ and ρ_ξ on the expected firm profit is driven by the value of the processing option of the firm. Since the firm optimally

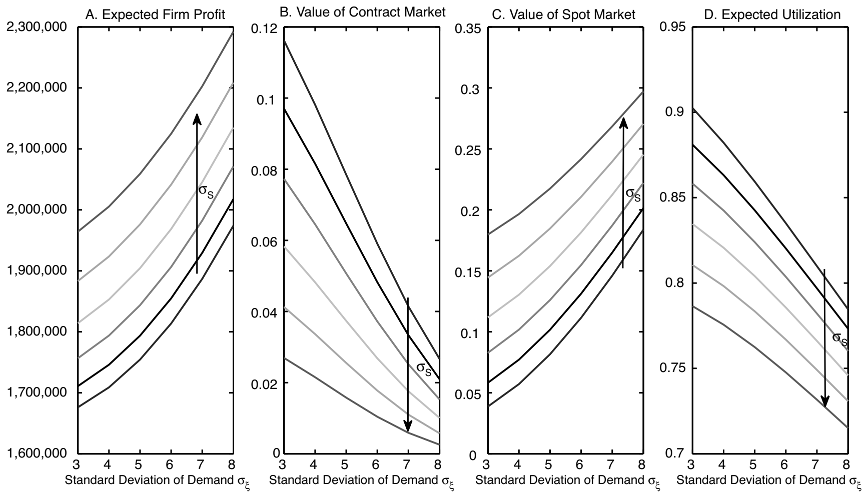


FIGURE 17.5 Impact of spot price variability (σ_S) and product market variability (σ_ξ) on the expected firm profit, value of contract, and spot market and expected utilization: σ_S ranges from 4% to 9% of the mean spot price (μ_S) with 1% increments and σ_ξ ranges from 3% to 8% of the mean demand of the standard product (μ_2) with 1% increments.

processes only when product market return is sufficiently high, a higher variability of product market return (i.e. a higher σ_ξ or ρ_ξ, increases the value of processing option of the packer, and thus, the expected optimal profit. The result with respect to σ_ξ is depicted in panel A of Figure 17.5.

With an increase in σ_S or σ_ξ, a lower (higher) dependence on contract (spot) market leads to a lower (higher) value of contract (spot) market as depicted in panel B (C) of Figure 17.5. The reduction in the volume of C-cattle processing dominates the increase in the volume of S-cattle processing and the expected total number of input processed decreases. As a result, expected utilization decreases (panel D). These results continue to hold with an increase in ρ_ξ.

17.4.2 EFFECT OF CONTRACT MARKET TRANSACTION COSTS (v AND ω)

In this section, we analyze the effect of the transaction cost for spot sales (ω) and the value surcharge for the quality difference of C-cattle (v) on the key performance indicators. As ω increases, the value of spot resale of the C-cattle decreases. As v increases, the contract procurement cost increases. Therefore, as depicted in panel A of Figure 17.6 below, with an increase in ω or v, the optimal contract volume decreases. In turn, the expected spot procurement increases (panel B) and the optimal portfolio ratio decreases (panel C).

The increase in the contract procurement cost (with an increase in v) and the decrease in the profitability of spot resale (with an increase in ω) decreases the

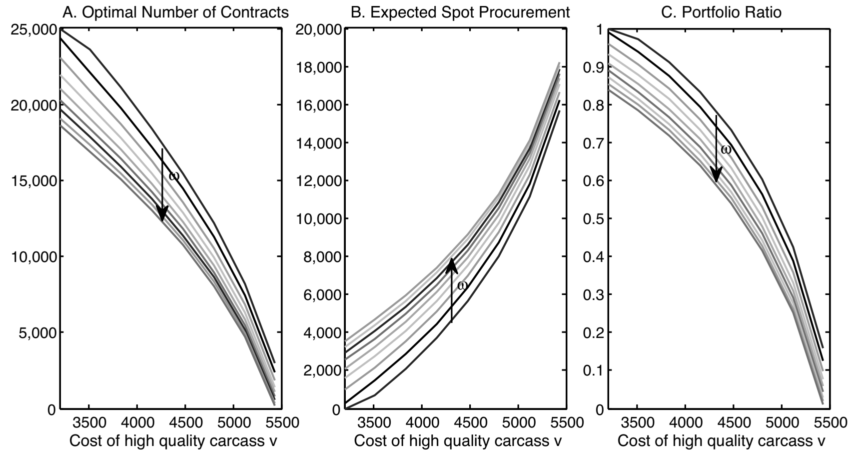


FIGURE 17.6 **Impact of transaction cost in spot sales (ω) and surcharge for high quality carcass (v) on the optimal procurement portfolio: ω ranges from 0% to 4% with 0.5% increments and v ranges from 3200 to 5440 with 320 increments (or equivalently, the quality premium $v\Delta$ ranges from 2.5% to 4.25% of the mean spot price μ_S with 0.25% increments).**

expected firm profit as depicted in panel A of Figure 17.7. With an increase in v or ω, a lower (higher) dependence on the contract (spot) market leads to a lower (higher) value of the contract (spot) market as observed in panel B (panel C).

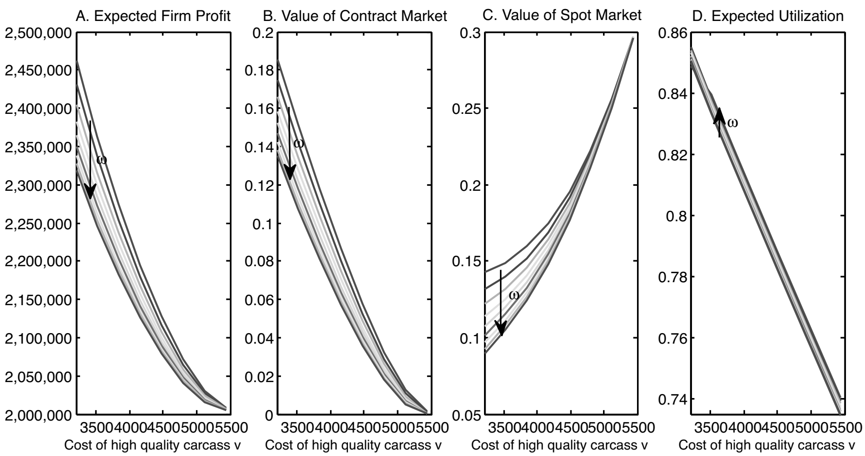


FIGURE 17.7 **Impact of transaction cost in spot sales (ω) and surcharge for high quality carcass (v) on the expected firm profit, value of contract and spot market, and expected utilization: ω ranges from 0% to 4% with 0.5% increments and v ranges from 3200 to 5440 with 320 increments (or equivalently, the quality premium $v\Delta$ ranges from 2.5% to 4.25% of the mean spot price μ_S with 0.25% increments).**

The decrease in the volume of C-cattle processing outweighs the increase in the volume of S-cattle processing and the expected utilization decreases (panel D).

17.4.3 EFFECT OF QUALITY DIFFERENCE BETWEEN C-CATTLE AND S-CATTLE (Δ)

As Δ increases, there are two opposite effects, the cost effect and the revenue effect. On the cost side, the contract procurement cost increases as the additional surcharge is tied to Δ. On the revenue side, the premium (standard) product yield from C-cattle increases (decreases). Consistent with the practice, in our numerical experiments, we observe that the premium product market is more profitable than the standard product market. Therefore, a higher Δ increases the value of C-cattle processing.

As depicted in panel A of Figure 17.8, with an increase in Δ, the cost effect dominates the revenue effect and Q^{C*} decreases. For a given Q^C, expected spot procurement is independent of Δ. Since Q^{C*} decreases, the expected spot procurement increases (panel B) and the optimal portfolio ratio decreases (panel C). It is interesting to note that even when there is no quality difference ($\Delta = 0$), the packer optimally contracts up to full capacity K. Despite the early commitment requirement of contract procurement, additional non-uniformity processing cost δ and transaction cost t of S-cattle together with the low level spot resale transaction cost ω induce the packer to prefer C-cattle over S-cattle.

For the effect on the expected profit, the cost effect dominates the revenue effect and the expected profit decreases with an increase in Δ as depicted in panel A of Figure 17.9. A lower (higher) dependence on the contract (spot) market leads to a lower (higher) value of the contract (spot) market (panel B). The expected

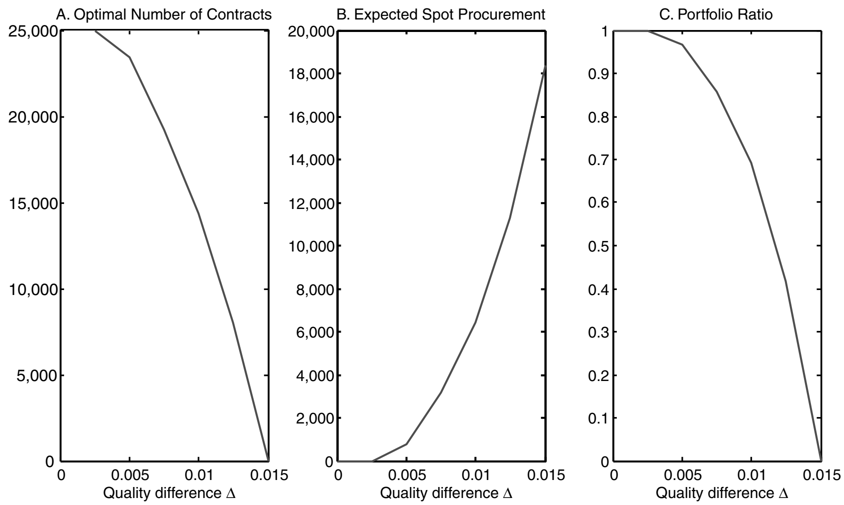


FIGURE 17.8 Impact of quality difference (Δ) on the optimal procurement portfolio: Δ ranges from 0 to 0.015 with 0.0025 increments.

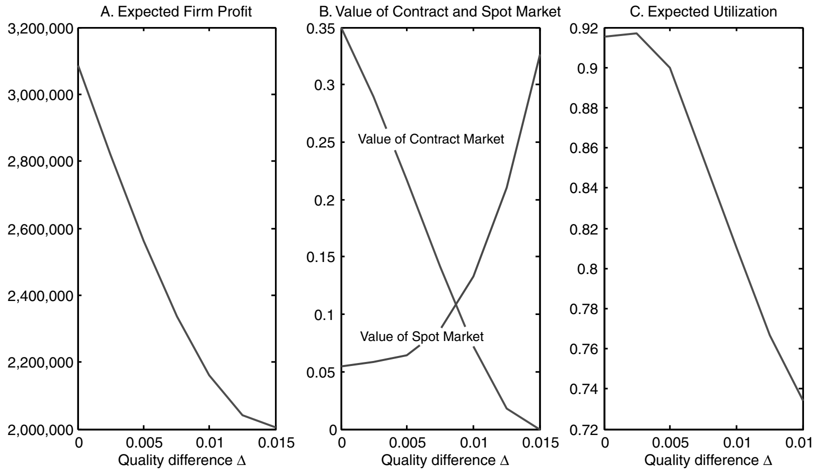


FIGURE 17.9 Impact of quality difference (Δ) on the expected firm profit, value of contract and spot market, and expected utilization: Δ ranges from 0 to 0.015 with 0.0025 increments.

utilization first increases then decreases as shown in panel C. For significantly low levels of Δ, the packer optimally contracts up to the full plant capacity and there is no spot procurement. In this case, an increase in Δ increases the value of C-cattle processing and a lower volume of C-cattle is sold to the spot market. Therefore, the expected utilization increases. For higher levels of Δ, the packer contracts less than the plant capacity and relies on the spot procurement. In this case, with an increase in Δ, the decrease in the volume of C-cattle processing outweighs the increase in the volume of S-cattle processing and the expected utilization decreases.

17.4.4 EFFECT OF UTILIZATION COST PARAMETER C_1

As depicted in panel A of Figure 17.10, a higher c_1 increases the optimal volume of C-cattle: As the cost of underutilization of the plant capacity K increases, the packer contracts more to lessen the impact of underutilization. In other words, the contract market provides a hedge against increasing utilization penalty cost. Although for a given Q^C the expected spot procurement would increase for the same reason, a higher Q^{C*} decreases the expected spot procurement. Therefore, the optimal portfolio ratio increases (panel C).

A higher c_1 decreases the expected profit as depicted in panel A of Figure 17.11. Since the firm relies more (less) on the contract (spot) market with an increase in c_1, the value of the contract (spot) market increases (decreases) as shown in panel B (panel C). Since the packer almost never sells back the C-cattle to the spot market, and uses C-cattle for processing; with an increase in c_1, the increase in the volume of processed C-cattle outweighs the decrease in the volume of S-cattle and the expected utilization increases as depicted in panel D.

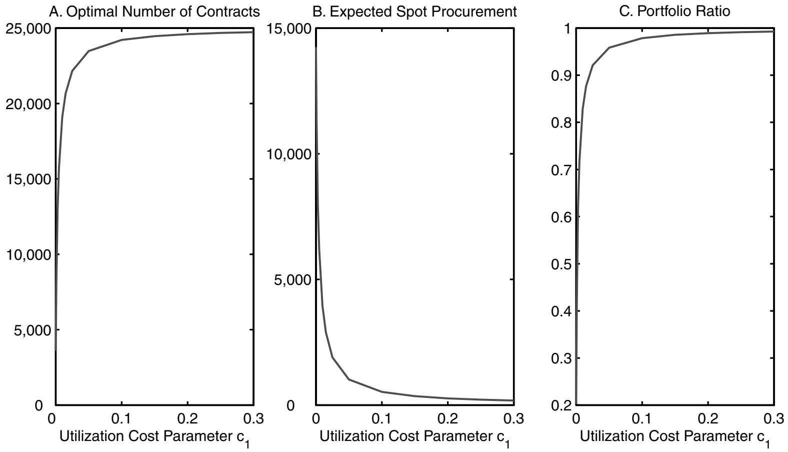


FIGURE 17.10 **Impact of utilization cost parameter (c_1) on the optimal procurement portfolio: c_1 range is in the set of** $\{0, 0.001, 0.003, 0.005, 0.01, 0.015, 0.025, 0.05, 0.1\}$.

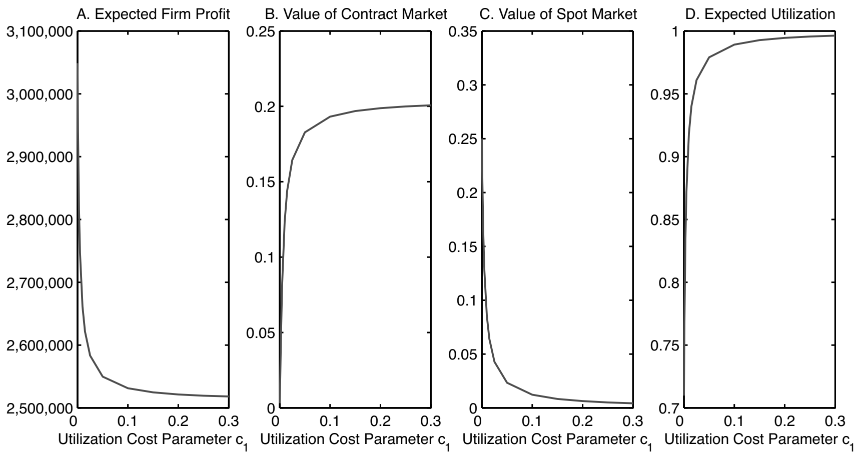


FIGURE 17.11 **Impact of utilization cost parameter (c_1) on the optimal procurement portfolio and expected firm profit: c_1 range is in the set of** $\{0, 0.001, 0.003, 0.005, 0.01, 0.015, 0.025, 0.05, 0.1\}$.

17.4.5 EFFECT OF DEMAND AND PRODUCT SUBSTITUTION

The effect of *demand substitution* (through the cross-price elasticity parameter e) is driven by the change in the product market profitability, and hence the value of processing. As e increases, since the two outputs are substitutes, for fixed production levels, the price of each product decreases. This leads to a lower product market profitability as the firm is not able to price differentiate between the two markets due to the higher cross-price effect. Therefore, higher demand substitution decreases the value of processing. It follows that Q^{C*} decreases with an

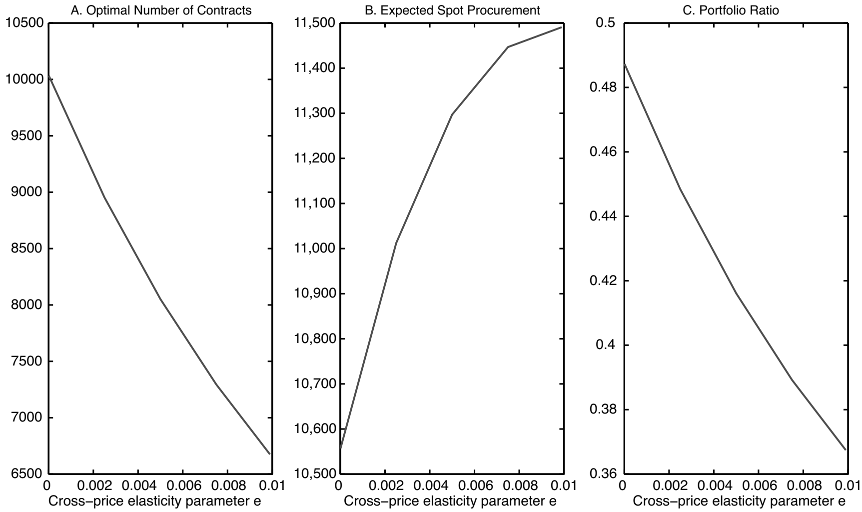


FIGURE 17.12 Impact of demand substitution (through cross-price elasticity parameter *e*) on the optimal procurement portfolio: *e* ranges from 0 to 0.01 with 0.0025 increments.

increase in e, as depicted in panel A of Figure 17.12. Although for a given Q^C, the expected spot procurement decreases in e due to lower value of processing, the reduction in Q^{C*} leads to an increase in the expected spot procurement (panel B). Therefore the optimal portfolio ratio decreases (panel C).

As depicted in panel A of Figure 17.13, with an increase in the cross-price elasticity parameter e, a lower value of processing decreases the expected firm profit. A lower (higher) dependence on the contract (spot) procurement leads to a lower (higher) value of the contract (spot) market as depicted in panel B. The decrease in the volume of C-cattle processing outweighs the increase in the volume of S-cattle processing and the expected utilization decreases (panel C).

The effect of *product substitution* is driven by the product substitution regime used by the firm. To understand the extent of product substitution, we explicitly calculate the expected premium product substitution ratio $\frac{\mathbb{E}[x_{12}^*]}{\mathbb{E}[x_{11}^*+x_{12}^*]}$ in our numerical experiments. However, product substitution does not have any value for the calibration implied by the GIPSA data; for this data the firm optimally does not use any product substitution. This observation is consistent with empirical observations, as packers rarely convert premium product (program beef) to standard product (commodity beef) in practice.

We note here that the ineffectiveness of product substitution partly depends on the high value of product market correlation ρ_ξ. The optimal substitution regime is determined by the difference between two market prospects. As demand correlation decreases, the asymmetry between the two markets increases, and the firm starts using partial and full product substitution regimes. As depicted in Figure 17.14 below, the expected premium product substitution ratio increases with a decrease in ρ_ξ for sufficiently negative correlation levels. In this case, product substitution does have a significant effect on the key performance measures.

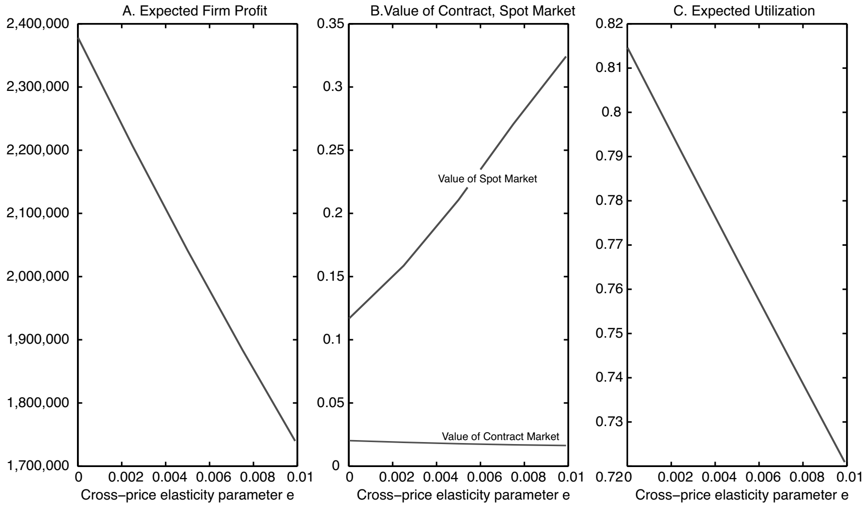


FIGURE 17.13 Impact of demand substitution (through cross-price elasticity parameter *e*) on expected firm profit, value of contract and spot market and expected utilization: *e* ranges from 0 to 0.01 with 0.0025 increments.

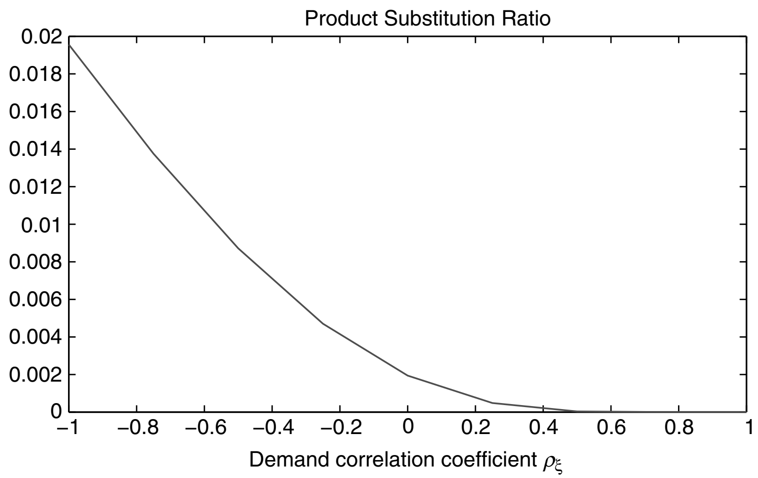


FIGURE 17.14 Impact of demand correlation (ρ_ξ) on the optimal expected Product 1 substitution ratio $\left(\frac{\mathbb{E}[x_{12}^*]}{\mathbb{E}[x_{11}^*+x_{12}^*]}\right)$: ρ_ξ ranges from -1 to 1 with 0.25 increments.

17.5 Discussion

Our results provide insights on several open questions of importance to the beef industry, including the efficiency and value of contract markets, which has been a fundamental bone of contention in the beef industry for decades. Among others, we have the following managerial insights. Lower variability in the input and output markets increases the value of the contract market relative to the spot market. Thus, the packer should increase the contract procurement with a decrease in variability. Interestingly, the packer does not benefit from lower variability. This is because the packer makes money out of the uncertainties in the market place (both spot and product market). Since the firm optimally responds to such uncertainties, lower variability in the markets decreases expected profits (in the usual spirit of real options). A higher quality difference between fed cattle sourced from contract and spot market does not necessarily benefit the packer, as this difference is reflected in the surcharge premium of the contract price. When the packer faces an increase in the utilization penalty costs, the contract market should be used more extensively to secure processing volume to hedge against the increasing processing costs.

It is important to bear in mind that the calibration for the numerical studies reported was undertaken at mean values of the parameters reported for the period October 2002 to March 2005. For this base case, the value of the contract market was not high (see Table 17.4). However, there were significant periods during the time frame of the GIPSA study in which the input and output market parameters dictated a much higher value of contract markets, as our sensitivity analysis indicated (e.g., the impact of σ_S and σ_ξ). Indeed, central to understanding the value of the contract market for packers is the variability in market parameters across time and the relative fixedness of packer technology and cost structures. The flexibility accorded by increased sourcing alternatives, including the contract market, is therefore extremely important in responding to market fluctuations over the life of the packer's plant.

The usual caveats apply in interpreting the results of a single set of parameters. Even with this caveat in mind, what is apparent in the present context is the richness of the interactions across various drivers of the key performance indicators. One of the most important elements of the beef context is the fact that, as is typical in fed-cattle markets, contract prices and spot prices are closely linked through the standard contract. Even with this close link, the sensitivity of the optimal portfolio to variability in both upstream and downstream markets is significant (e.g., see Table 17.4). What this indicates is a strong interaction among upstream and downstream factors. This is all the more evident when considering the impact on optimal contracts, profits and utilization from the other factors characterizing these markets. For example, changes in quality determinants of the contract (captured in Δ) can have significant impacts on the optimal portfolio. Of course, the main drivers of the optimal portfolio are the mean values of prices of contract and spot cattle, and the price sensitivity and variability in the final product markets. All of these vary considerably over time depending on supply and demand of the respective cattle entering into these two markets (e.g., See Figure 2.1 in the GIPSA Report [2007] and the ensuing discussion, which describes very significant changes over time in prices in the U.S. beef industry during the period 2002–2005 of that study). As a result, what one can expect is that the optimal portfolio, and the value of the contract market itself, will change over time, and at times dramatically, as determinants of supply-demand and prices change. This is consistent with the basic story of this paper and other contributions to supply management under risk: Namely, there is real value in the integration of risk management, production and marketing, and all the more so under conditions of varying environmental conditions and fixed plant size and technology.

There are a number of limitations to the present study. The model analyzed reflects the specific characteristics of the U.S. beef market, which has a number of idiosyncrasies, including the pricing of contract procurement relative to the spot market. In other contexts, the price in the contract purchases could well be fixed and/or subject to other determining factors (e.g., the competitive model developed in Wu and Kleindorfer [2005]). Moreover, even for other live animal supply chains, such as pork-hog and broiler-chicken, there are important differences from the beef market For example, for the pork-hog market, one would see a higher proportion of the premium product (i.e., $a_1 > a_2$), in contrast to the beef supply chain, and the optimal operating regime would be different with important consequences for different substitution results. These comments and noted limitations suggest a number of open research questions.

There are several empirical avenues that are opened by the results of this study. These include both comparisons of different size plants, and of the performance and structure of sourcing portfolios as market conditions vary. In addition to these matters of direct interest to both industry and policy makers, there are also other interesting features in the model presented that deserve empirical study. These include the effect of contracting terms (such as options and resale value parameters), utilization and scale effects (which are reported to be extremely important in packer decisions), and the impact of price level and volatility on spot and contract cattle purchasing decisions. These are all very interesting for the beef industry. In addition, other effects modeled here, such as product and demand substitution, may be even more important in other markets.

Concerning risk management, our focus has been on physical procurement only. Extensions to overlay the cash flows from this physical problem with financial hedging are an important area of future research. In the beef industry, for example, there are significant variations over time in market conditions and operating profits of meat packers. To the extent that profit smoothing would avoid financial transactions costs under such variable market conditions, financial hedging can be of significant value. Financial options defined on either input or output markets can serve this purpose. As noted in Kleindorfer (2008), these hedge markets need not be identical with the sourcing markets as long as they are sufficiently highly correlated with these markets.

In addition to short-term issues, there are also important capacity investment and technology choice issues in the longer term. Intuitively, it is clear that the trade-offs involved between scale economies, operational flexibility (in downward substitution and yields) are likely to be richer and more complex in a fixed proportions technology world than in a single-input, single-output world. From the numerical analysis in this paper, we already see that these trade-offs will involve complex interactions between the magnitude of the scale economies and the entire fabric of the short-term optimization problem (solved here for the beef market) given capacities. A deeper examination of these with an appropriate temporal separation between capacity/technology choices and shorter term operating and contracting choices would be interesting.

REFERENCES

Antonovitz F., Green R. 1990. Alternative estimates of fed beef supply response to risk. *American Journal of Agricultural Economics*, **72**:475–487.

Azzam A. 1998. Captive supplies, market conduct, and the open-market price. *American Journal of Agricultural Economics*, **80**:76–83.

Bitran G.R., S.M. Gilbert. 1994. Co-production processes with random yields in the semiconductor industry. *Operations Research*, **42**:476–491.

Boetel B.L., Liu D.J. 2003. Evaluating the effect of generic advertising and food health information within a meat demand system. *Agribusiness: An International Journal*, **19**:345–354.

Boyabatlı O., P.R. Kleindorfer, S.R. Koontz. 2011. Integrating long-term and short-term contracting in beef supply chains. *Forthcoming in Management Science*.

Braschler C. 1983. The changing demand structure for pork and beef in the 1970s: Implications for the 1980s. *Southern Journal of Agricultural Economics*, **15**:105–110.

Brester G.W., J.M. Marsh. 2001. The effects of U.S. meatpacing and livestock production technologies on marketing margins and prices. *Journal of Agricultural and Resource Economics*. **26**:445–462.

Chavas J.P. 1983. Structural change in the demand for meat. *American Journal of Agricultural Economics*, **65**:148–153.

Crespi J.M., Sexton R.J. 2004. Bidding for cattle in the texas panhandle. *American Journal of Agricultural Economics*, **86**:660–674.

Dahlgran R.A. 1987. Complete flexibility systems and the stationarity of U.S. meat demand. *Western Journal of Agricultural Economics*, **12**:152–163.

Deaton A. Muellbauer J. 1980. An almost ideal demand system. *American Economic Review*, **70**:312–326.

Hayenga, M., T. Schroeder, J. Lawrence, D. Hayes, T. Wukina, C. Ward, W. Purcell. 2000. Meat packer vertical integration and contract linkages in the beef and pork industries: An economic pespective. American Meat Institute Report, Iowa.

Hennesy D.A. 2005. Slaughter house rules: Animal uniformity and regulation for food safety in meat packing. *American Journal of Agricultural Economics*, **87**:600–609.

Kastens T.L., Brester G.W. 1996. Model selection and forecasting ability of theory-constrained food demand systems. *American Journal of Agricultural Economics*, **78**:301–312.

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

Kleindorfer P.R. 2008. Integrating physical and financial risk management in supply management. Forthcoming in: Geman H. editor. *Risk Management for Commodity Markets*. New York: Wiley Finance.

Koontz S.R., J.D. Lawrence. 2010. Impacts of alternative marketing agreement cattle procurement on packer costs, gross margins, and profits: Evidence from plant-level profit and loss data. *Agribusiness*. **26**:1–24.

Kuchler F., McClelland J. 1989. Issues raised by new agricultural technologies: Livestock growth hormones. *Agricultural Economic Report*, 608. USDA-ERS.

Burton D.M., Love H.A. 1999. A strategic rationale for captive supplies. *Journal of Agricultural and Resource Economics*, **24**:1–18.

MacDonald J.M. 2003. Beef and pork packing industries. In S.R. Koontz (ed.). The veterinary clinics of North America Food Animal Practice: Economies of the Red Meat and Diary Indutries, *Saunders*, 419–443.

Marsh J.M. 1983. A rational distributed lag model of quarterly live cattle prices. *American Journal of Agricultural Economics*, **65**:539–547.

Marsh J.M. 1984. Estimating slaughter supply response for U.S. cattle and hogs. *North Central Journal of Agricultural Economics*, **6**:18–28.

Marsh J.M. 1994. Estimating intertemporal supply response in the fed beef market. *American Journal of Agricultural Economics*, **76**:444–453.

Marsh J.M. 2003. Impacts of declining U.S. retail beef demand on farm-level beef prices and production. *American Journal of Agricultural Economics*, **85**:902–913.

Moschini G., Meilke K.D. 1989. Modeling the pattern of change in U.S. meat demand. *American Journal of Agricultural Economics*, **71**:253–261.

Mullen J.D., Wohlgenant M.K., Farris D.E. 1988. Input substitution and the distribution of surplus gains from lower U.S. beef-processing costs. *American Journal of Agricultural Economics*, **70**:245–254.

Nelson G., Spreen T. 1978. Monthly steer and heifer supply. *American Journal of Agricultural Economics*, **60**:117–125.

GIPSA (Grain Inspection Packers and Stockyards Administration) Report. 2007. GIPSA livestock and meat marketing study; vol 3: Fed cattle and beef industries. Washington D.C.: U.S. Department of Agriculture (USDA).

Reutlinger S. 1966. Short run beef supply response. *American Journal of Agricultural Economics*, **48**:909–919.

Schmitz J.D. 1997. Dynamics of beef cow herd size: An inventory approach. *American Journal of Agricultural Economics*, **79**:532–542.

Schroeder J.R., Azzam A. 2003. Captive supplies and cash market prices for fed cattle: The plant-level relationship. *Agribusiness: An International Journal*, **19**:489–504.

Schroeder J.R., Azzam A. 2004. Captive supplies andthe spot market price of fed cattle: The role of delivery timing incentives. *Agribusiness: An International Journal*, **20**:347–362.

USDA (U.S. Department of Agriculture) 2009. Agricultural Statistics 2009. USDA National Agricultural Statistics Service. Washington.

Verbeke W., Ward R.W. 2001. A fresh meat almost ideal demand system incorporating negative TV press and advertising impact. *Agricultural Economics*, **25**:359–374.

Wang Y., Jaenicke E.C. 2006. Simulating the impacts of contract supplies in a spot market-contract market equilibrium setting. *American Journal of Agricultural Economics*, **88**:1062–1077.

Ward C.E., T.C. Schroeder. 2002. Structural changes in cattle feeding and meat packing. In *Managing for Today's Cattle Market and Beyond*. Western Extension Marketing Committee, Colorado.

Wu D.J., P.R. Kleindorfer. 2005. Competitive options, supply contracting and electronic markets. *Management Science*, **51**:452–466.

Xia T., Sexton R.J. 2004. The competitive implications of top-of-the-market and related contract-pricing clauses. *American Journal of Agricultural Economics*, **86**:124–138.

Zhang M., Sexton R.J. 2000. Captive supplies and the cash market price: A spatial markets approach. *Journal of Agricultural and Resource Economics*, **25**:88–108.

Zhang M., Sexton R.J. 2001. FOB or uniform delivered prices: Strategic choice and welfare effects. *Journal of Industrial Economics*, **49**:197–221.

CHAPTER EIGHTEEN

Risk Management in Electric Utilities

STEIN-ERIK FLETEN, JUSSI KEPPO, AND ERKKA NÄSÄKKÄLÄ

18.1 Introduction

Utility companies minimize the fluctuation of their cash flows by identifying, monitoring and hedging different risk factors. In this chapter we analyze volume and price risks, and discuss how they can be modeled and hedged. As in many financial markets, the electricity market participants face liquidity risk, credit risk, operational risk, and political risk, and we discuss shortly about these as well.

In order to understand electricity price fluctuations, electricity supply and demand has to be analyzed over time. Seasonal variations in the supply and demand cause cycles in a spot price. For example, in the mountain areas of Norway

Acknowledgments: We thank TrønderEnergi for providing data. Fleten acknowledges financial support from the Research Council of Norway through project 178374/S30. All errors are ours.

Handbook of Integrated Risk Management in Global Supply Chains, First Edition. Edited by Panos Kouvelis, Lingxiu Dong, Onur Boyabatli, and Rong Li.

the precipitation during winter accumulates as snow and, therefore, the electricity supply from hydro reservoirs is higher during spring and summer when the snow melts, whereas the demand there has an opposite seasonality. The cold winter days in Norway cause an extra heating load to the system. Supply-demand equilibrium changes also due to the changes in business cycle, regulation, and politics. For instance, an aluminium plant scaling down production due to the lower demand of its products and a coal fired power plant closing down due to increased emission costs change the market equilibrium. All the examples above give typical sources of electricity price uncertainty that utility companies have to manage.

Similarly as with the price risk, many electricity market participants have to also manage volume risk continuously since electricity consumption and production processes may change rapidly and depend, at least partly, on factors that are not controlled by the market participants. For example, in the case of wind power production the total electricity production varies as a function of the weather, and the production of a hydropower plant depends partly on uncertain inflow. This means that producers do not exactly know the size of their future production, and hence they face the volume risk. This holds also with consumption since the consumption fluctuations depend mostly on the weather. That is, electricity consumption cycles can be explained by weather changes. This also implies that the short-term price elasticity is almost zero.

Energy companies use electricity derivative instruments in their risk management. An electricity derivative is a financial contract whose value depends on the electricity spot price. For example, a forward contract is an obligation to buy or sell electricity for a predetermined price at a predetermined future time. When companies use derivative contracts, they face a risk that the number of counterparties willing to enter the other side of the transaction can change drastically. Therefore, the price might change even if the underlying spot price and other fundamentals do not change. Further, if the used hedging strategy is leveraged and liquidity vanishes from the credit market then there might be a funding problem with the hedging strategy. These types of risks are called liquidity risk. For example, during the global financial crisis of 2008–2009 most banks drastically scaled down their lending and trading, including commodity trading. This led to a significantly lower liquidity (and higher premiums) in the electricity derivative markets.

Market participants do not face counterparty risk if the transactions are cleared in organized exchanges. This is because clearinghouses of the exchanges collect margins from the counterparties guaranteeing that winners get their gains and losers pay their losses. On the other hand, if a transaction is not cleared in an exchange and if credit risk mitigation methods, such as netting, collaterization and downgrade triggers are not used, then the transaction most likely has counterparty risk.

The usage of power exchanges and their clearinghouses varies in different countries. For instance, in Europe the percentage of volume traded on the exchanges compared to total electricity consumption ranges from 0.7% (France) to 29.7% (Nordic countries: Denmark, Finland, Iceland, Norway and Sweden). These transactions can be viewed as credit risk-free and the rest of the transactions might have some counterparty risk. Thus, credit risk management is an important topic in electricity markets since most of the transactions are done outside the exchanges.

According to the Basel Committee (2001), operational risk is the risk of loss resulting from inadequate or failed internal processes, people, and systems, or from external events. This definition includes, for example, people risks, technology and processing risks, physical risks, and legal risks, but it excludes reputation risk and strategic risk. Due to the Basel Committee, operational risk management usually considers financial firms. However, the operational risk management is important also in utility companies as the Northeast Blackout of 2003 illustrated. This blackout affected about 10 million people in the Canadian province of Ontario and about 45 million people in eight U.S. states. The main causes of the blackout were one energy company's failure to trim trees in part of its Ohio service area and a computer bug that prevented the company from warning other companies about the blackout. These led to cascading failures in the electricity grid (see Andersson et al. 2005).

Political risk is especially prominent in electricity markets, since transition from fossil fuels to alternative technologies is needed to reduce greenhouse gases. For instance, in Europe the transition is sought through the EU-wide emission trading scheme and through country-specific support for renewable energy generation. The combination of these two policy instruments is likely to dampen the growth of both carbon and electricity prices, and hamper development of radically new technologies.

The rest of the chapter is organized as follows. Section 18.2 considers electricity price risk and Section 18.3 volume risk. Section 18.4 discusses other risks, and finally Section 18.5 concludes.

18.2 Price Risk

Electricity companies use similar risk measures as financial companies. The most common risk measure is the value at risk (VaR) that gives the maximum loss under a certain confidence level. The time horizon in the calculations depend on the liquidity of the portfolio since the unwinding of a portfolio depends on the liquidity of the assets. For instance, if the portfolio consists only liquid tradable assets then a 10 day VaR is convenient. For highly illiquid portfolios a significantly longer time is used (e.g., one year).

Many of the electricity companies' assets are not traded. These include power plants and inventories. However, their values can be estimated by using simple models that depend on electricity and raw material forward prices and/or on their expected future prices. Due to the used models, the price estimates of nontraded assets most likely include model risk. In order to understand the magnitude of this risk, several model candidates should be used.

Also other risk measures can be used, for instance, expected shortfall, which is the expected loss given that the loss is greater than the VaR level. For more about different risk measures see McNeil et al. (2005). In order to calculate the risk measures the main risk factors have to be identified. In electricity markets typically these are price and volume risks. We first discuss the price risk which means the risks in electricity spot and forward prices.

18.2.1 ELECTRICITY FORWARD CURVE

Forward contracts are the simplest and most common derivatives in the electricity markets. Together with electricity spot price (starting point of the forward curve) they are viewed as the underlying assets for other derivatives, and embody the main revenue risk factors of electricity market participants. Therefore, the modeling of electricity forward and spot prices is vital for many analyses in this market.

Let us consider a small example motivating the importance of the forward prices as a source of future price estimates when electricity production is valued. A company owns 16 MWh of weekly electricity production capacity. The next week's production is sold on an electricity market. The company estimates that the next week's spot price is 50 €/MWh with 50% probability and 32 €/MWh with 50% probability. On the other hand, in the market the next week's forward price for electricity is 40 €/MWh. Note that there is no arbitrage even though $40 \neq 50\% * 50 + 50\% * 32$, because electricity cannot be stored. The forward prices are the market's view on the expected future spot price and its risks. That is, a forward price is the expected spot price under so-called risk-neutral pricing measure and this can be different from the expected spot price under the objective probability measure. Note that the value of the next week's production under the risk-neutral measure is €640 and under the company's objective measure it is €656. Thus, if the company wants to hedge its production plan with forward contracts, it gets €640 without risk, and if it does not hedge then the the revenues are random: €512 with 50% probability and €800 with 50% probability.

When a forward contract is traded, there is a market participant willing to buy electricity at a specified future date with the price dictated by the forward contract. Correspondingly, there is also a market participant that is willing to sell electricity with the same price in the future. This means that the forward prices reflect the expected future electricity supply-demand equilibrium and risks involved in the equilibrium, i.e., the forward prices can be seen as risk adjusted expected future spot prices. The time to the delivery of the electricity is called maturity. The different maturity electricity forwards form an electricity forward curve. There are cycles and peaks in the forward curve due to the seasonality in the supply and demand of electricity. The starting point of the forward curve approaches the electricity spot price as the forward maturity goes to zero. Figure 18.1 illustrates the electricity forward curve in the Nordic countries on August 11, 2009.

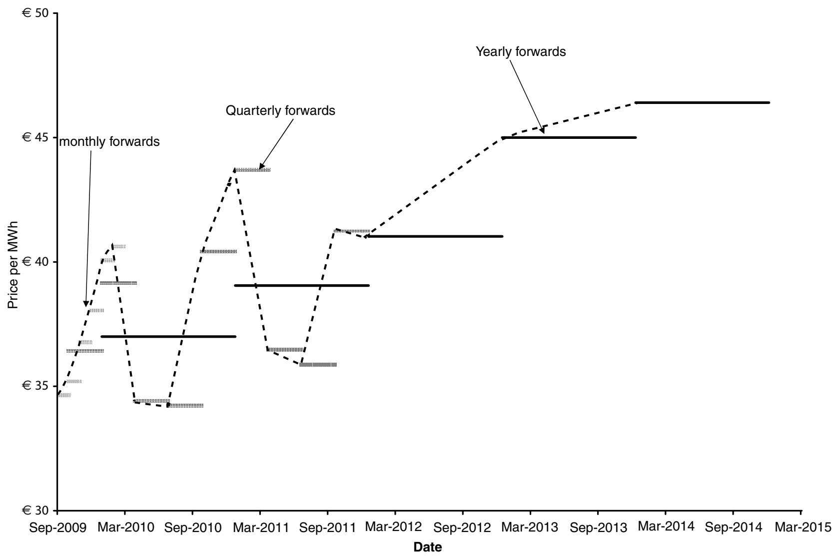


FIGURE 18.1 **Nord Pool's (Nordic Power Exchange) forward curve on August 11, 2009.**

The currency used at Nord Pool (marketplace for power in the Nordic countries) is Euro, so their energy prices are in €/MWh. There are weekly, monthly, quarterly, and yearly forwards. The delivery period (i.e., the duration of the forward contracts) increases as a function of the maturity. Weekly contracts are traded for the nearest 4–7 weeks, monthly contracts for the nearest 6 months, quarterly contracts for the nearest 8-11 quarters, and yearly contracts for the nearest 5 years. The forward curve in Figure 18.1 illustrates that the market expects electricity prices to increase in December and January and then decrease towards summer. Due to the cold winter in the Nordic countries, the prices in January–April are traditionally higher than in summer and autumn. The Nord Pool's forward curve gives information about the expected weekly variations only for the nearest 4–7 weeks (this is not shown in Figure 18.1). In many cases forward curve with weekly granularity is needed over a longer period than these 4–7 weeks. For example, Fleten and Lemming (2003) present a method to estimate a weekly forward curve based on longer-term forwards and forecasts generated by bottom-up models.

Björk and Landen (2002), Miltersen and Schwartz (1998), Clewlow and Strickland (1999), among others, consider the modeling of electricity forward curve dynamics. Koekebakker and Ollmar (2005) and Audet et al. (2004) report that the key characteristics of the electricity forward curve are:

- Spot volatility curve
- Forward volatility curve's maturity effect
- Forward curve's correlation structure

The spot volatility curve describes seasonal variations in the spot price uncertainty. For instance, in the Nordic countries the highest volatility is from spring to fall since then hydropower systems face inflow uncertainty. Due to the mean-reversion in the spot price, forwards with long time to maturity tend to vary less than forwards with a shorter maturity. Thus, the forward volatility falls in the time to maturity and this is called the maturity effect. Forwards with maturity dates close to each other correlate more than forwards that have delivery dates far from each other. This is a typical correlation structure between commodity forwards.

One way to model the above discussed characteristics is to use the following price process for the electricity forward contracts:

$$dS(t, T) = \exp(-\alpha(T-t))\sigma_S(T)S(t,T)dB_T(t) \quad \forall \quad t \in [0, T] \qquad (18.1)$$

In (18.1) $S(t,T)$ is T-maturity forward price at time t (T is fixed), α is a strictly positive constant, $\sigma_S(T)$ is deterministic spot volatility at time T, and $B_T(t)$ is a standard Brownian motion corresponding to the T-maturity forward price. The correlation between two forward prices, $S(t, T_1)$ and $S(t, T_2)$, are modeled as $\exp(-\beta|T_1 - T_2|)$, where $T_1$and T_2 are the maturities of the forward contracts, $\beta > 0$ is the rate the correlation falls as a function of the maturity difference.

Note that in (18.1) T is fixed, so we model $S(t+dt, T) - S(t, T)$. In the limit $t \to T$ we have $S(T, T)$ which is the spot price at time T. Equation (18.1) implies that the forward volatility is lower than the corresponding spot volatility. The parameter α models the exponential decrease in the forward volatility as a function of maturity. The decrease in the forward volatility can be seen as a consequence of the mean reverting nature of electricity spot prices (see, e.g., Clewlow and Strickland 1999). Equation (18.1) is used (e.g., in Audet et al. 2004, Koekebakker and Ollmar 2005). Eydeland and Wolyniec (2003) summarize several possible price processes for electricity in the U.S. markets, and their empirical investigation provide mixed results about those processes. Given the information at time t, from equation (18.1) we get that the spot price at time T (i.e. $S(T) = S(T, T)$) is log-normally distributed with mean $S(t, T)$ and:

$$Var(S(T)) = \frac{S^2(t, T)\sigma_S^2(T)}{2\alpha}\left[1 - \exp\left(-2\alpha(T-t)\right)\right]. \qquad (18.2)$$

By (18.1), the forward price is a martingale and, therefore, it would be natural to assume that $B_T(t)$ there is a Wiener process under so-called risk neutral pricing measure Q (see, e.g., Duffie 2001). However, in this chapter we mostly assume that the risk-neutral pricing measure equals the objective probability measure P (see more discussion on this in subsection 18.3.1). Thus, the dynamics in (18.1) can be viewed also as the observed forward price process. Note, however, that in general there is no reason to assume that the forward price is a martingale under the objective measure.

18.3 Volume Risk

Volume hedging reduces the future cash flow fluctuation due to the changes in volume. One way to do this is to find a portfolio of standard derivative instruments that replicates the part of the future cash flows that depend on the volume. Selling this portfolio provides the hedge. This kind of portfolio selection problem in complete markets is studied, for example, in Cox and Huang (1989), Karatzas et al. (1987), Cvitanic and Karatzas (1996) and Shreve and Soner (1994). These models consider continuous time hedging. In reality transaction costs and illiquidity can make the use of dynamic hedging strategies expensive and difficult. Carr et al. (1998), and Carr andWu (2002), among others, study static portfolios of standard derivatives replicating the payoffs of a given derivative. These hedging strategies are static as the optimal weights of the derivatives in the portfolio are constant. Hence, the replicating portfolio is not adjusted dynamically. When static hedging strategies are used the transaction costs are lower and the implementation of the hedging strategies is easier. For example, by Cvitanic et al. (1999), it is not optimal to hedge continuously European-type contingent claims with proportional transaction costs. Instead a discrete time hedging or even buy and hold strategy should be implemented.

From the viewpoint of a market participant one major difference between electricity market and stock market is that electricity consumption/production is given by an exogenous process while a stock investor can decide asset holding himself. For example, in the case of hydropower production the total electricity production depends on the inflow to the hydro reservoirs, which depends on the amount of rainfall. Often the uncertainty in the production/consumption process does not perfectly correlate with the electricity derivative prices. This means that the volume uncertainty cannot be fully hedged with those derivatives. In practice, there is always some volume risk that the producers and consumers cannot hedge.

Portfolio/risk management in electricity markets has been studied, for example, in Pilipovic (2007), Fleten et al. (2002; 2009a,b), Aïd (2009) and Vehviläinen and Keppo (2003). For the particular case of hydropower production, the results in Fleten et al. (2009c) are useful in connection with hedging. The authors find that production in week t, $p(t)$ depends on the size of the plant D_{cap} (equals one for big power plants and otherwise it's zero), the hydro runoff to the reservoir in the current week $w(t)$, the spot price $S(t, t)$ relative to forward price $S(t, T)$ (here we use the average of forward prices for next week and next quarter) and on last week's production $p(t-1)$ as follows:

$$p(t) = -1045.25 + 913.35D_{cap} + 0.07w(t) - 0.05D_s w(t) + 1361.90\frac{S(t, t)}{S(t, T)} + 0.87p(t-1) + \epsilon(t)$$

Note that in the above equation all the parameters are statistically significant at the 5% level, $\epsilon(t)$ is a normally distributed error term, and D_s is a filling season indicator (equals one during the filling season and otherwise it's zero)

TABLE 18.1 The Probabilities of Different Production-Price Scenarios

Production	32 €/MWh	50 €/MWh
12 MWh	10%	40%
20 MWh	40%	10%

that indicates that the inflow affects production differently at different times (inside/outside the filling season). The forward price $S(t, T)$ here is the average price of the nearest week forward (having maturity next week) and the nearest quarter forward, maturing the next quarter. The out-of-sample R^2 is 88%, indicating that this model explains most of the variation in the production volume. By (18.3), the volume is characterized relatively accurately, and the explicit dependence on spot and forward prices makes the regression model useful for hedging purposes.

Let us continue our simple example in Section 18.2. There we assumed that the size of the production is constant. Now let us assume that also the production has uncertainty which partly correlates with the price uncertainty. More precisely, there are two possible outcomes for the production and two possible outcomes for the spot price (i.e., there are four possible production-price scenarios for the next week). The probabilities of different scenarios are given in Table 18.1. The production and price have a negative correlation, that is, when the price is low (32 €/MWh), the production is 12 MWh with 20% probability and 20 MWh with probability 80%, and when the price is high (50 €/MWh), the production is 12 MWh with 80% probability and 20 MWh with probability 20%. Note that in this case the unconditional expected value of the production is 16 MWh (high/low prices have the same probability), which is equal to the deterministic production in the previous case.

The next week's forward price is again 40 €/MWh. In Figure 18.2 the probabilities of different portfolio values, when the production is partially hedged by selling 16 MWh of forwards, are illustrated. As there is also uncertainty in the production process the price hedging does not totally remove the uncertainties in the future cash flows. The corresponding future value of the unhedged portfolio varies between €384–1000, whereas the future cash flows of the price hedged portfolio vary between €440–840. The standard deviation of the unhedged portfolio is €142 and the standard deviation of the price hedged portfolio is €167. Hence, in this example hedging using the expected production value increases the standard deviation of the portfolio, and in this sense it cannot be optimal. However, the hedging strategy has increased the worst possible outcome from €384 to €440. This illustrates how the uncertainty in the production process complicates the hedging decisions considerably.

In the next subsection we will introduce a simple static hedging strategy for electricity production/consumption.

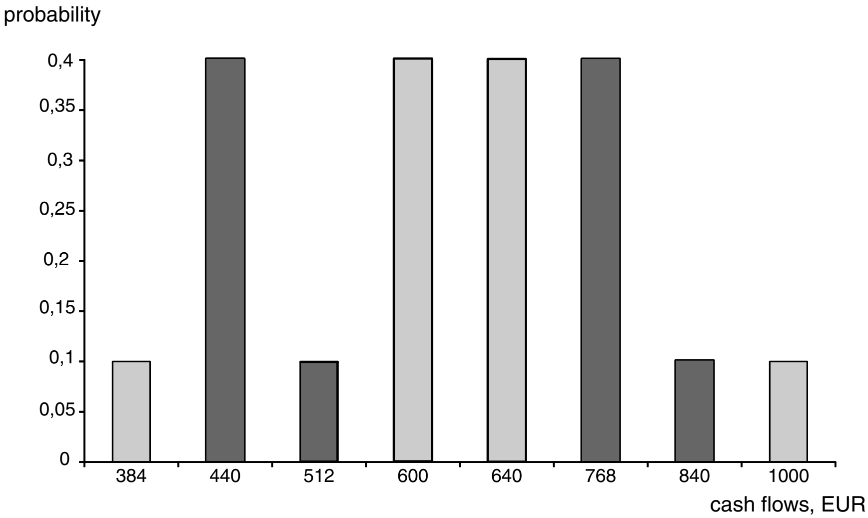


FIGURE 18.2 **Probabilities of different portfolio values. The black bars are the probabilities for the case when the production is hedged by selling 16 MWh of forwards, and the grey bars are for a portfolio that is not hedged.**

18.3.1 STATIC HEDGING

In this subsection we follow Näsäkkälä and Keppo (2005). We consider an electricity market where spot and derivative instruments are traded continuously in a finite time horizon. We assume that there exist forward contracts on electricity spot price, and that the electricity derivative market is complete and there is no arbitrage. The no arbitrage assumption states that all portfolios with the same future payoffs have the same current value. The no-arbitrage condition and the completeness of the market ensure the existence of a unique linear pricing function. The linear pricing function can be described by the risk-neutral pricing measure Q. Under Q all the expected returns of traded non-dividend paying financial assets are equal to the risk-free interest rate r (see, e.g., Duffie, 2001). Thus, at time t the price of T-maturity derivative on electricity spot price is given as

$$f(t, T) = \exp(-r(T-t))E_t^Q[\phi(S(T, T))] \quad \forall\, t \in [0, T] \tag{18.3}$$

The payoff function is $\phi(\cdot)$, $S(T, T)$ is the electricity spot price at time T, E_t^Q is the expectation operator under the risk-neutral probability measure Q and with respect to the information at time t. For simplicity, risk-free rate r is assumed to be constant. Note that since electricity is not a financial asset its expected return under Q is not usually equal to r, as is the case for all tradable non-dividend paying assets.

As mentioned above, we denote the T-maturity forward price at time t by $S(t, T)$. By allowing T to vary from t to $\tilde{T}$ we get the forward curve

$S(t,\cdot):[t,\tilde{T}]\rightarrow R_{+}$, where the maximum maturity $\tilde{T}>t$. There are cycles and peaks in the forward curve due to the seasonality in electricity spot price. The starting point of the forward curve $S(t,t)$ is the electricity spot price at time t. Because the value of a forward contract is by definition zero when initiated, we get from (18.3):[1]

$$S(t,T)=E_t^Q[S(T,T)]=E_t[M(T)S(T,T)] \tag{18.4}$$

$M(t)$ is Radon-Nikodym derivative dQ/dP and E is the expectation operator under the objective measure P. For simplicity, we assume here that $M(t)=1$ for all t and, thus, the objective measure equals the risk-neutral pricing measure. The forward dynamics are introduced in (18.1).

We consider consumption/production at time $T>t$. Let $V(t,T)$ denote the T-maturity volume estimate of the consumption/production at time t, that is:

$$V(t,T)=E_t[v(T)] \tag{18.5}$$

Note that $v(T)$ is the volume at time T. We assume that the process of T-maturity volume estimate follows

$$dV(t,T)=V(t,T)\left[\sigma_V(t,T)dB_V(t)+\sigma_{VS}(t,T)dB_T(t)\right] \tag{18.6}$$

In (18.6) $\sigma_V(t,T)$ and $\sigma_{VS}(t,T)$ are bounded deterministic volatility functions. B_T is the Brownian motion corresponding to T-maturity forward contract and B_V is a standard Brownian motion independent of B_T. Thus, uncertainty in the T-maturity volume estimate due to the changes in electricity forward prices is modeled with B_T, while B_V models the uncertainty that is uncorrelated with the corresponding forward price. Therefore, at time t the volatility of T-maturity volume estimate is $\sqrt{\sigma_V^2(t,T)+\sigma_{VS}^2(t,T)}$ and the correlation with T-maturity forward price is:

$$\rho(t,T)=\frac{\sigma_{VS}(t,T)}{\sqrt{\sigma_V^2(t,T)+\sigma_{VS}^2(t,T)}} \tag{18.7}$$

The volume process (18.5) gives a log-normally distributed volume with mean $V(t,T)$ and variance:

$$Var(v(T))=V^2(t,T)\left[\exp\left(\int_t^T\left[\sigma_V^2(s,T)+\sigma_{VS}^2(s,T)\right]ds\right)-1\right], \tag{18.8}$$

where $v(T)$ is the volume at time T. Note that as we model the conditional expectations there are no restrictions on the volume pattern as a function of maturity.

An agent able to adjust the electricity volume due to adverse changes in forward prices has a negative correlation (i.e., $\sigma_{VS}(t,T)<0$). For example, an

[1] The payoff of the forward at time T is $S(T,T)-S(t,T)$ and the value of the forward at t is zero. Thus, by setting $f(t,T)=0$ and $\phi(S(T,T))=S(T,T)-S(t,T)$ in (18.3), we get (18.4).

electricity consumer can decrease its electricity consumption when the forward prices are high. The adjustments can be done, for example, by changing the performance level of an electricity consumption unit. The consumption process parameters can be estimated, for example, by using Räsänen et al. (1995, 1997). An agent with a negative correlation is called a flexible consumer. Correspondingly, an agent able to exploit positive changes in the forward prices has a positive correlation, i.e., $\sigma_{VS}(t, T) > 0$. For example, an electricity producer can increase its production by starting up flexible production units that are used only when electricity prices are high. The optimal production plan can be solved, for example, by using Gjelsvik et al. (1992), Pereira (1989), and Keppo (2002). An agent with a positive correlation is called as a flexible producer. Finally, an agent with zero correlation is called as a nonflexible agent. When there is no correlation, the total volatility is $\sigma_V(t, T)$, i.e., $\sigma_{VS}(t, T) = 0$.

We solve the optimal hedging strategy of T-maturity electricity production/consumption by minimizing the variance of the future cash flows. If the cash flow distribution is approximately normally distributed then this can be viewed as minimizing the VaR of the cash flows since in this case VaR is linear with respect to the standard deviation of the cash flows.

We assume that the optimal hedging strategy is given by optimal hedging amount and time. We solve the hedging amount in terms of the hedge ratio that describes the hedging size as a proportion of the volume estimate. We assume that after the hedging time the position is not readjusted even if there are remarkable changes in the volume estimate or in the forward prices. Thus, before the hedging time the agent makes each moment a decision whether to hedge with the current volume estimate or wait and make the decision in the future. The electricity cash flows of an agent at time T are given by

$$\pi(\tau, T, \eta(\tau)) = V(T, T)S(T, T) + \eta(\tau)V(\tau, T)\left[S(\tau, T) - S(T, T)\right] \quad (18.9)$$

Note that $\tau \in [0, T]$ is the hedging time, $\eta(t)$ is the hedge ratio and, therefore, $\eta(\tau)V(\tau, T)$ is the number of forwards in the portfolio. The above described hedging strategy is a buy and hold strategy. The strategy is easy to implement, transaction costs are small and they are known in advance. In contrast to static strategies, the transaction costs of dynamic hedging strategies are often high and they depend on the amount of readjustments done on the portfolio. Moreover, static hedging strategies are more usable in illiquid markets, because continuous trading are often too costly (see, e.g., Carr and Wu, 2002). Note that the static forward hedging strategies can remove only linear dependencies on the forward prices and the dynamic strategies hedge perfectly if the volume uncertainty depends only on the forward prices.

Let us consider a numerical example. The time horizon is one year (i.e., $T = 1$). The flexible producer faces positively correlated forward price and production estimate. In this case we assume that the correlation is 0.5. Table 18.2 summarizes the production process parameters and the optimal hedging strategy. The optimal hedging time is obtained by minimizing the variance of the cash flows in (18.9).

TABLE 18.2 Flexible Producer's Production Process Parameters and the Corresponding Optimal Hedging Strategy

	ρ	σ_V	σ_{VS}	τ	η
Value	0.50	0.87	0.05	0.00	1.21

TABLE 18.3 Flexible Consumer's Production Process Parameters and the Corresponding Optimal Hedging Strategy

	ρ	σ_V	σ_{VS}	τ	η
Value	−0.50	0.87	−0.05	0.57	0.82

That is:

$$\min_{\tau \in [0, T], \eta \in (-\infty, \infty)} Var\left[\pi\left(\tau, T, \eta\right)\right]$$

The hedge ratio η (as a function of τ) can be solved solved analytically from the first order conditions. However, the hedging time τ has to be solved numerically. In this flexible producer case the hedging time is 0, thus it is optimal to hedge the portfolio immediately. As the optimal hedge ratio is 1.21, it is optimal to over-hedge the portfolio. This over-hedging is due to the positive correlation between $V(T, T)$ and $S(T, T)$ in (18.9).

The consumption of the flexible consumer correlates negatively with the electricity prices. We assume that the correlation is -0.5. The consumption process parameters and the hedging strategy are summarized in Table 18.3. The optimal hedging time is 0.57, thus it is optimal to hedge 5 months before the maturity. The optimal hedge ratio is 0.82 (i.e., in this case it is optimal to under-hedge the portfolio). The under-hedging is caused by the negative correlation between $V(T, T)$ and $S(T, T)$ in (18.9), (i.e., the negative correlation creates a partial hedge and, thus, the flexible consumer does not hedge that much).

The volume estimate of the nonflexible agent does not correlate with the forward price. In this case the optimal hedging time is 0.17, thus it is optimal to hedge the portfolio 10 months before the maturity. The nonflexible agent's optimal hedge ratio is always one. The parameters and the hedging strategy are summarized in Table 18.4.

In Figure 18.3 the standard deviation of the portfolio process is given as a function of the hedging time for all the three cases. The standard deviation curves have the same volume volatility, but the correlations are different. The

TABLE 18.4 Non-Flexible Agent's Production/Consumption Process Parameters and the Corresponding Optimal Hedging Strategy

	ρ	σ_V	σ_{VS}	τ	η
Value	0	0.10	0.00	0.17	1

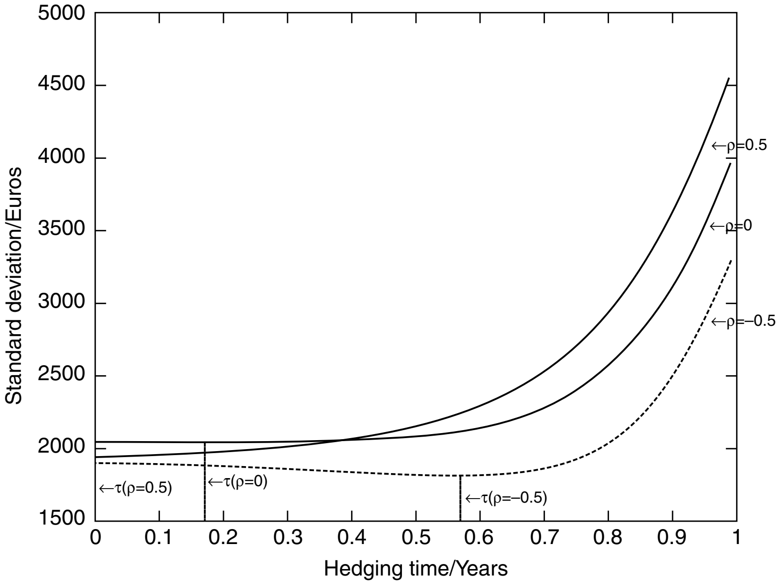


FIGURE 18.3 Standard deviations as a function of hedging time for typical market participants. The forward price $S(0, 1) = 20$ euro/MWh and the load estimate is 1 GWh. The flexible producer is the gray line, while the dashed black line is the flexible consumer and the black line is the nonflexible agent.

smaller the correlation the later the standard deviation attains its minimum. Thus, positive correlation makes early hedging more favorable while negative correlation postpones the hedging decision. Negative correlation can be seen as an additional way to hedge against price changes. Note that changes in correlation change also the minimum value of the standard deviation. For example, as shown in Figure 18.3 the portfolio with negative correlation can be hedged most effectively.

In Figure 18.4 the hedge ratios are presented for all the three cases. The figure indicates that when the correlation is positive it is optimal to over-hedge, and when the correlation is negative it is optimal to under-hedge. Further, when the correlation is zero, the optimal hedging amount is equal to the volume estimate. The optimal hedge ratios get closer to one as the maturity decreases. Hence, negative correlation can be seen as a hedge against price changes whereas the positive correlation is an additional source of risk.

18.4 Other Risk Examples

In this section we shortly discuss some other risks that affect electricity market participants.

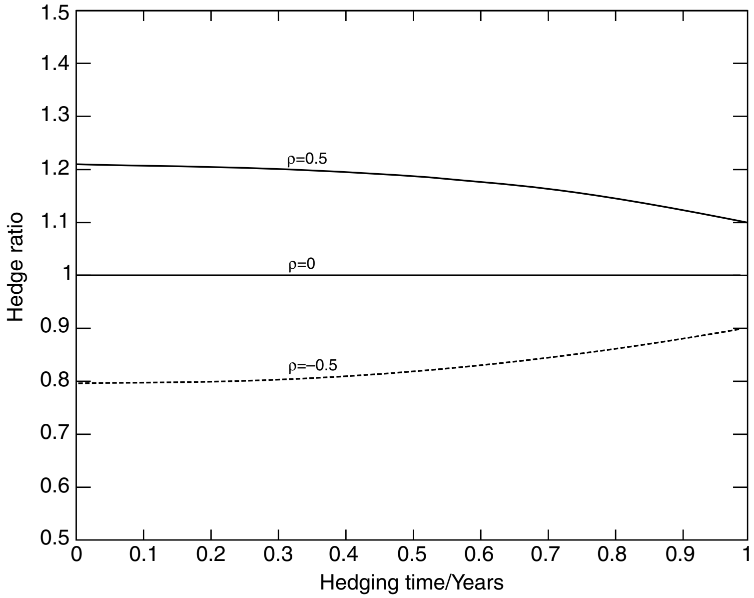


FIGURE 18.4 Optimal hedge ratios for typical market participants. The forward price $S(0, 1) = 20$ euro/MWh and the load estimate is 1 GWh. The flexible producer is the gray line, while the dashed black line is the flexible consumer and the black line is the nonflexible agent.

18.4.1 LIQUIDITY RISK

Let us follow the ideas in Jonsson et al. (2004). We consider a energy company (or a trading firm) that acts in a market consisting of a risk-free asset and risky electricity forward contracts. Because the derivative market might be illiquid, the forward prices depend on the actions of the energy company. This is clear when bid-ask spreads are wide. A nonzero bid-ask spread can be modeled by using pricing and hedging frameworks with transaction costs (see, e.g., Clewlow and Hodges 1997, Hodges and Neuberger 1989, Keppo and Peura 1999, and Musiela and Rutkowski 1998).

We model the realized forward prices as a function of two variables: the quantity that the energy company trades in the market and the effect from other market participants. The effect from the other market participants is modeled with risk factor processes that follow geometric Brownian motion processes with time dependent volatility as in Section 18.2:

$$\frac{ds(t, T)}{s(t, T)} = \sigma(t, T)\, dW_T(t) \tag{18.10}$$

Note that $s(t, T)$ is the value of the factor at time t for T-maturity forward contract, $\sigma(t, T)$ is the time dependent (deterministic) volatility, and W_T is Brownian motion for the T-maturity forward price. If s goes up (down) then the demand relative to the supply rises (falls) and the forward price goes up (down). The price impact due to the energy company's actions is modeled as follows

$$F(t, T, q) = I(q, T)s(t, T) \tag{18.11}$$

In (18.11) q is the quantity of T-maturity forwards hold by the energy company and $I(\cdot, T)$ is the price impact for the T-maturity forward contract describing the highest (lowest) selling (buying) price if the quantity q is bought (sold), $F(t, T, q)$ is the T-maturity forward price at time t under quantity q. We assume $I(q, T)$ is bounded and positive for all q and T, $I(q, T) > 1$ if $q > 0$, $I(q, T) < 1$ if $q < 0$, $I(0, T) = 1$ for all T, and $I(q, 0) = 1$ for all q. That is, the smaller the trading position and the closer to the maturity, the smaller the price impact. Thus, there is no price impact on the spot price. According to previous empirical studies (Loeb 1983, Kraus and Stoll 1972, Holthausen et al. 1987, Keim and Madhavan 1996), the price impact depends on the market capitalization and the liquidity of the market.

The liquidity could vanish also from the credit market. This can be modeled, for example, by a stochastic maximum debt level as follows. The maximum leverage level (here the leverage is measured as the debt level relative to the portfolio value) has two states: high (L_h) and low (L_l). That is, the fraction of the energy company's debt to the value of its total assets is bounded by L_l in the low state and by L_h in the high state. The change of the state is modeled as follows:

$$\begin{aligned}
P(L(t+dt) = L_h | L(t) = L_l) &= \lambda_l dt \\
P(L(t+dt) = L_l | L(t) = L_l) &= 1 - \lambda_l dt \\
P(L(t+dt) = L_l | L(t) = L_h) &= \lambda_h dt \\
P(L(t+dt) = L_h | L(t) = L_h) &= 1 - \lambda_h dt
\end{aligned}$$

Note that λ_l and λ_h are constant, and they represent the speed of state changes (from l to h and vice versa).

Interest rates are mostly driven by macro factors, not the actions of the energy company. Therefore, the risk-free asset is given by:

$$\frac{dB(t)}{B(t)} = r(t)\,dt$$

Note that $B(t)$ is the risk-free asset, r is the instantaneous borrowing/lending rate and we assume here that it follows a mean-reverting dynamics:

$$dr(t) = (b - ar(t))dt + \sigma_r dW_r(t)$$

The parameters a, b, and σ_r are constant but could depend on the state (high/low). Parameter a is the speed of mean reversion and fraction b/a is the mean level. Parameter σ_r is the interest rate volatility. For instance, when σ_r is high then the borrowing/lending rate fluctuates a lot.

Let us define the liquidation function as $G(h, T) = \int_0^h I(-q, T)dq$ where h is the number of T-maturity forward contracts currently held by the energy company. If the energy company holds h forwards at time t then the liquidation value at that time equals $G(h, T)s(t, T)$ and this is the value the energy company gets when it sells the forward contracts at time t.

We define the company's total wealth in the financial instruments to be equal to the money it receives after liquidation of all its assets, that is:

$$V(t) = m(t)B(t) + \sum_{i \in \{1,\ldots,n\}} G(h_i(t), T_i)s(t, T_i)$$

The number of different maturity forward contracts in the portfolio is n, h_i is the holding for T_i-maturity forwards, m is the amount of money in the risk-free asset (if $m < 0$ then borrowing) and, by the borrowing constraint, $-m(t)/V(t) \leq L_h$ in high state and $-m(t)/V(t) \leq L_l$ in low state. If the energy company trades in a self-financing way then we have:

$$dV(t) = m(t)dB(t) + \sum_{i \in \{1,\ldots,n\}} G(h_i(t), T_i)ds(t, T_i) \tag{18.12}$$

An interesting interpretation of equation (18.12) is that the wealth dynamics of a self-financing strategy $(m, h_1, \ldots, h_n)$ is the same as that of an agent applying the self-financing strategy $(m, G(h_1, T_1), \ldots, G(h_n, T_n))$ in a market without price impacts and where the forward price processes are given by the risk factors (18.10). Therefore, by (18.10) to (18.12), we can analyze the illiquid market similarly as in Sections 18.2 and 18.3 and, due to the shape of $I(\cdot, T)$, the electricity company hedges less in quantity sense, because the hedging affects the forward prices.

18.4.2 OPERATIONAL AND POLITICAL RISKS

The operational risk in electric utilities is connected to trading of the financial products and to the physical power plants. Many companies have an experience of managing operational risk through procedures and contingency planning, and through the use of insurance covering damage to generation or transmission/distribution assets caused by (e.g., natural disasters, fire, liability regarding damage to a third party, and insurance against crimes and vandalism). These plans include IT-systems risk with respect to availability of equipment, and information security regarding internal databases and analysis tools. They further include health, safety and environmental risks through procedures in the major activities in all business units. For hydropower companies, dam safety is an area of special focus. Another is the potential damage to fish population as a consequence of intense ramping of hydropower in rivers and dammed lakes. For nuclear plants, safety needs to be built into the culture of the company.

Examples of operational risk incidents include rouge trading. This happened in the Oslo-based electricity utility E-CO in a case that surfaced December 2008. A trader had taken positions far beyond his limits in financial instruments, and the CEO Hans Erik Horn had to leave the company immediately. There was no evidence indicating that the trader had done this for personal gain. Losses were limited however, and did not affect the earnings significantly.

Another example of operational risk is the unforeseen outage in Trollheim power plant owned by Statkraft in Norway, July 2007. The outage caused the Surna river flow to drop dramatically, which in turn killed between 500 000 and 1 million salmon fry. Statkraft agreed to pay 9 million NOK (1.6 million USD) to the local river owner organization and to the municipality in September 2008.

Political risks affect the business conditions of electric utilities. This includes general taxes and electricity-specific taxes, environmental taxation, and taxation of electricity consumers. There might be changes in the conditions in operating licenses as a consequence of political decisions. Finally, many utilities are owned by municipalities, counties and states, meaning that (e.g., their dividend and debt policy is determined by politicians).

An example of political risk affecting electricity companies in the Nordic countries is the potential increase in the number of price zones. Today, Sweden and Finland are two separate price zones, while Denmark and Norway have several zones within their borders. Electricity spot prices differ across zones, reflecting differences in supply, demand and transmission capacities. If, for example, Sweden is split into two zones, it is likely that producers in the north will receive permanently lower prices, while producers in the south will enjoy permanently higher spot prices. Prices in neighboring zones will also be affected, for example, producers in Mid-Norway will likely experience lower prices. An increase in the number of price zones is discussed among the energy ministers in the Nordic countries, aiming for implementation in 2010.

18.5 Summary

In this chapter we have discussed risk factors that affect electricity market participants. We introduced price and volume risks and a simple model to hedge them. We also discussed liquidity, operational, and political risks that are more difficult to quantify (e.g., due to the lack of data). However, since they have significant impact on the cash flows, the market participants have to consider these factors, for instance, by scenario analysis.

REFERENCES

Aïd R. 2009. Long-term risk management for utility companies: The next challenges. Working paper. Laboratoire de Finance des Marchés d'Energie, Université Paris Dauphine—Paris IX-CREST-EdF R& D, August.

Andersson G. et al. 2005. Causes of the 2003 major grid blackouts in North America and Europe, and recommended means to improve system dynamic performance. *IEEE Transactions on Power Systems*, 20(4):1922–1928.

Audet N., Heiskanen P., Keppo J., Vehviläinen I. 2004. Modeling electricity forward curve dynamics in the Nordic market. In: Bunn D.W., editor. *Modelling Prices in Competitive Electricity Markets*. New York: Wiley Financial.

Basel Committee on Banking Supervision. 2001. *Sound Practices for the Management and Supervision of Operational Risk*. December, Basel, Switzerland.

Björk T., Landen C. 2002. On the term structure of forward and futures prices. In: *Mathematical Finance Bachelier Congress 2000*. Springer Verlag.

Carr P., Wu L. 2002. Static hedging of standard options. *CRIF Working Paper series*, New York.

Carr P., Ellis K., Gupta V. 1998. Static hedging of exotic options. *Journal of Finance*, 1165–1190.

Clewlow L., Hodges S. 1997. Optimal delta-hedging under transactions costs. *Journal of Economic Dynamics and Control*, 21:1353–1376.

Clewlow L., Strickland C., 1999. Valuing energy options in a one factor model fitted to forward prices. Quantitative Finance Research Group, Working paper. Sydney, Australia: University of Technology.

Cox J.C., Huang C.F. 1989. Optimal consumption and portfolio policies when asset prices follow a diffusion process. *Journal of Economic Theory*, 49(1): 33–83.

Cvitanic J., Karatzas I. 1996. Hedging and portfolio optimization under transaction costs: A martingale approach. *Mathematical Finance*, 6(2):133–165.

Cvitanic J., Pham H., Touzi N. 1999. Super-replication in stochastic volatility models under portfolio constraints. *Journal of Applied Probability*, 36(2): 523–545.

Duffie D. 2001. *Dynamic Asset Pricing Theory*. Princeton, New Jersey: Princeton University Press.

Eydeland A., Wolyniec K. 2003. *Energy and Power Risk Management*. Hoboken, New Jersey, Wiley.

Fleten S.-E., Lemming J. 2003. Constructing forward price curves in electricity markets. *Energy Economics*, 25(5):409–424.

Fleten S.-E., Wallace W.S., Ziemba W.T. 2002. Hedging electricity portfolios using stochastic programming. In: Greengard C., Ruszczyński A., editors. *Decision Making under Uncertainty: Energy and Power*, volume 128 of *IMA Volumes on Mathematics and Its Applications*. New York: Springer-Verlag. pages 71–93.

Fleten S.-E., Bråthen E., Nissen-Meyer S.-E. 2009a. Evaluation of static hedging strategies for hydropower producers in the Nordic market. Working paper. Norwegian University of Science and Technology, Trondheim, Norway.

Fleten S.-E., Bråthen E., Nissen-Meyer S.-E. 2009b. Evaluation of a dynamic hedging strategy for hydropower producers in the Nordic market. Working paper. Norwegian University of Science and Technology, Trondheim, Norway.

Fleten S.-E., Keppo J., Lumb H.K., Weiss V.K. 2009c. Hydro scheduling powered by derivatives. Working paper. Norwegian University of Science and Technology, Trondheim, Norway.

Gjelsvik A., Røtting T.A., Røynstrand J. 1992. Long-term scheduling of hydrothermal power systems. In: Broch E., Lysne D.K., editors. Proceedings of the 2nd International Conference on Hydropower Development (Hydropower92), Lillehammer, Norway. Balkema, Rotterdam, pages 539–546.

Hodges S.D., Neuberger A. 1989. Optimal replication of contingent claims under transactions costs. *The Review of Futures Markets*, 8:222–239.

Holthausen R., Leftwich R., Mayers D. 1987. The effect of large block transactions on security prices: A cross-sectional analysis. *Journal of Financial Economics*, 19(2):237–267.

Jonsson M., Keppo J., Meng X. 2004. Option pricing with an exponential effect function. Working paper. Ann Arbor, Michigan: University of Michigan.

Karatzas I., Lehoczky J.P., Shreve S.E. 1987. Optimal portfolio and consumption decisions for a "small" investor on a finite horizon. *SI AM Journal on Control and Optimization*, 25(6): 1557–1586.

Keim D.B., Madhavan A. 1996. The upstairs market for large-block transactions: Analysis and measurement of price effects. *Review of Financial Studies*, 9(1):1–36.

Keppo J., Peura S. 1999. Optimal portfolio hedging with nonlinear derivatives and transaction cost. *Computational Economics*, 13:117–145.

Keppo J. 2002. Optimality with hydropower system. *IEEE Transactions Power Systems*, 17(3):583–589.

Koekebakker S., Ollmar F. 2005. Forward curve dynamics in the Nordic electricity market. *Financial Management*, 31(6):74–95.

Kraus A., Stoll H.R. 1972. Price impacts of block trading on the New York Stock Exchange. *Journal of Finance*, 27(3):569–588.

Loeb T.F. 1983. Trading cost: the critical link between investment information and results. *Financial Analysts Journal*, 39:39–43.

McNeil A.J., Frey R., Embrechts P. 2005. *Quantitative Risk Management: Concepts, Techniques, and Tools*. Princeton, New Jersey: Princeton University Press.

Miltersen K.R., Schwartz E.S. 1998. Pricing of options on commodity futures with stochastic term structures of convenience yields and interest rates. *Journal of Financial and Quantitative Analysis*, 33(1):33–59.

Musiela M., Rutkowski M. 1998. *Martingale Methods in Financing Modelling*, Springer, New York.

Näsäkkälä E. and Keppo J. 2005. Electricity load pattern hedging with static forward strategies. *Managerial Finance*, 31(6):116–137.

Pereira M.V.F. 1989. Optimal stochastic operations scheduling of large hydroelectric systems. *International Journal of Electrical Power & Energy Systems*, 11(3):161–169.

Pilipovic D. 2007. *Energy Risk: Valuing and Managing Energy Derivatives*. McGraw-Hill, New York.

Räsänen M., Ruusunen J., Hämäläinen RP. 1995. Customer level analysis of dynamic pricing experiments using consumption-pattern models. *Energy*, 20(9):897–906.

Räsänen M., Ruusunen J., Hämälainen R.P. 1997. Optimal tariff design under consumer self-selection. *Energy Economics*, 19(2):151–167.

Shreve S. Soner M. 1994. Optimal investment and consumption with transaction costs. *Annals of Applied Probability*, 4(3):609–692.

Vehviläinen I., Keppo J. 2003. Managing electricity market price risk. *European Journal of Operational Research*, 145(1):136–147.

Chapter Nineteen

Supply Chain Risk Management: A Perspective from Practice

COLIN KESSINGER AND JOE MCMORROW*

In a typical enterprise, sourcing and supply chain teams function as the boundary between their firm and its supply base. This position requires these teams to build and execute sourcing strategies in an environment where demand, supplier performance, pricing, and material availability are constantly changing. Commonly, these teams manage organization spend of anywhere from 30% to 70% of their firm's revenue. The magnitude of the dollars at stake in sourcing decisions means that even small miscalculations in price or quantity can have dramatic effects on a company's margins, top-line performance through lost sales and balance sheet through inventory. Double-digit percentage reductions in stock prices, 9-digit misses in revenues, and 10-digit inventory write-offs are all well documented and recurring events attributable to mismatches between supply and demand.

Outsourcing of manufacturing, logistics, and basic operational activities such as procurement, have left firms with little appreciation for the complexity and

*Member and past Chairman of the Supply Chain Risk Leadership Council (SCRLC).

Handbook of Integrated Risk Management in Global Supply Chains, First Edition. Edited by Panos Kouvelis, Lingxiu Dong, Onur Boyabatli, and Rong Li.

vulnerabilities that have been introduced into their supply chain. Incorporating lean methodologies into these complex supply chains has brought even greater exposure to supply disruptions. Managing the supply base is further complicated by the technical sophistication and collaboration required to quickly respond to changes in demand. Supply-demand mismatches can be driven by a wide variety of unpredictable events, from improperly forecasted sales volumes to natural disasters. Supply and demand uncertainty increases the time pressure on the supply chain and amplifies the risk of having invested too much or too little.

Still, most supply chain organizations continue to place the greatest emphasis on "static" measures such as price, quality, and logistics. Changes in these measures during the course of supplier negotiations are most easily mapped back to the bottom line. As a result, these metrics are easiest to track and evaluate and are common metrics to reward supplier performance. Not surprisingly, most supply chain and procurement organizations focus on and excel at price reduction programs and do less to prepare the firm for changes in the business environment, or that of their suppliers. This behavior results in situations where immediate and urgent action is required at a time when management has little ability to respond.

This chapter will explore these issues more thoroughly, offer a practical approach for assessing risk and resiliency in the supply chain, and share a case study from Cisco Systems' supply chain organization. The Cisco case describes how their supply chain organization has approached the challenge of understanding and building resiliency into the supply chain. The case also offers a real life situation, when Cisco and the rest of the world learned of the tragic 7.9 magnitude earthquake that struck in the Sichuan province of China, May 2008. Additionally, this chapter will close with what many of today's leading supply chain risk experts believe is necessary to improve supply chain resiliency—an International Standard for Business Resiliency.

19.1 Defining Supply Chain Risk Management

From the sole proprietorship to the largest enterprise, risk is an inherent part of doing business, and often, one that cannot be avoided. A look back at the natural disasters, environmental accidents, technology mishaps, recessions, and man-made crises over the past 50 years amply demonstrates that disruptive incidents will happen, and can have a significant adverse effect on business as usual. In the face of a substantial disruption, an effective risk mitigation approach can mean the difference between the demise and the survival of a business. The challenge for any successful business is to learn how to identify, mitigate, and manage risk; it is a challenge that requires more than the creation of a local disaster recovery plan. To be effective, supply chain organizations must acknowledge the importance of having a comprehensive and systematic process of prevention, preparedness, mitigation, response, and recovery.

One of the most critical aspects of addressing supply chain risk is having visibility and influence beyond the borders of your organization and into the supply chain. Supply chain risk management (SCRM) which can be defined as: "The practice of managing the risk of any factor or event that can materially

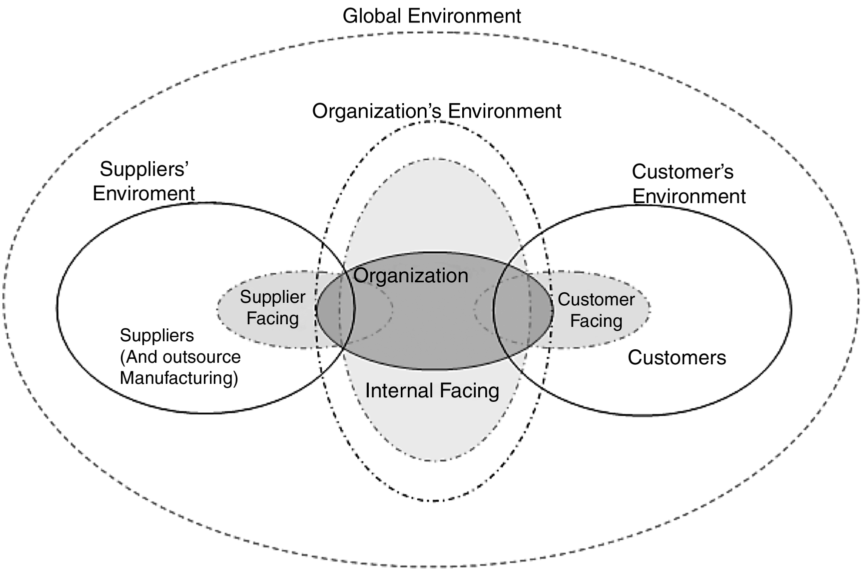


FIGURE 19.1 **Supply Chain Risk Environment.** ***Source:*** **SCC, Supply Chain Council.**

disrupt a supply chain, either within a single company or spread across multiple companies. The ultimate purpose of SCRM is to enable cost avoidance, continue to ensure customer service, and maintain market position."[1] According to a recent study from IBM titled "The Smarter Supply Chain of the Future," SCRM has emerged as the second largest challenge for supply chain executives after supply chain visibility—placing *even higher than increasing customer demands and higher costs*.[2] As a result, supply chains and the risks they face are now a prominent concern in the boardroom for many companies.

19.2 Understanding your Supply Chain

The Supply Chain Operations Reference Model (SCOR) was developed by management consulting firm PRTM and supply chain research firm, AMR Research, Inc. defines the core management processes within the supply chain as *plan, source, make, deliver,* and *return*. Risk management in the supply chain must first start with the evaluation of each of these core processes within the organization. Additionally, as referenced in the definition of SCRM, risk management in the supply chain must then extend beyond the internal processes and organizational boundaries. Managing risks in the supply chain requires an understanding of the organization's environment as well as the context of the global environment in which it operates, or that of the entire supply chain (Figure 19.1). Each piece

[1] IBM Global Chief Supply Chain Officer Study: The Smarter Supply Chain of the Future, February 2009

[2] Supply Chain Risk Leadership Council: Risk Lexicon, Andrew Cox, TSA, December 2008

of an organization's supply chain is exposed to risks, and each of its suppliers and customers must be engaged in some form of supply chain risk management for their organizations to ensure true end-to-end resiliency. Understanding your supply chain is a critical first step in determining the scope of an organization's SCRM program. Without this thorough understanding of the supply chain environment, an organization cannot adequately begin to identify, prioritize, and mitigate risks. In the following sections we talk more explicitly about the core capabilities associated with understanding our supply chain and the risk embedded in it.

19.3 Developing SCRM Capabilities

Despite the increase in visibility and importance of supply chain risk management among supply chain practitioners, business leaders, industry analysts, and the media, often even established organizations are embarking upon the creation of company-wide SCRM policies for the first time.

To successfully address this challenge, an organization must develop three core capabilities:

1. The ability to collect, communicate, and respond to new information as soon as it becomes available, both within the organization and in the supply chain
2. The ability to assess the impact of uncertainty about material requirements and supply performance on future sourcing performance.
3. The ability to quantify supply chain performance and design supply strategies to achieve sourcing objectives across uncertain business outcomes

19.3.1 CAPABILITY #1: COLLECT, COMMUNICATE, AND RESPOND

In the sourcing environment, critical planning and execution information flows include updated demand forecasts or production plans, information on inventory or material deliveries, new information about supplier capabilities or performance, and changes in technical requirements.

The infrastructure that enables firms to rapidly collect, communicate, and respond to new information includes planning, transaction, event management, and collaboration technologies. Most companies have already made substantial investments in both software and business processes to develop this capability to support their traditional business processes and mitigate risks such as minor supply disruptions (e.g., quality, line-down, etc.) and demand uncertainty. The benefit for most of these investments is reduced response times and shorter time to market with the right products. In essence, these capabilities enable the firm and its supply chain to respond more efficiently once uncertainty is resolved and key information is known. However, most companies have not extended their investment in these technologies to the full spectrum of risk, which includes more catastrophic events such as epidemics, geo-political uncertainty, or substantial swings in the economic cycles.

While this may or may not be where every company should start with an SCRM program, it often is the starting point with the least resistance. Most companies embrace the notion that they need to execute and respond more quickly and that these benefits extend to both regular operations as well as risk management activities.

19.3.2 CAPABILITY #2: ASSESSING THE IMPACT OF UNCERTAINTY

The presence of substantial uncertainty about the key drivers of sourcing performance—including material requirements, supply pricing and supply availability—drives equally large uncertainty about sourcing performance outcomes. As a result, managing the impact of these uncertainties can result in significant cost reductions, often in the 5% to 15% range. However, this also serves as proof of the old adage, "If you can't measure it, you can't manage it." Thus, creating visibility into the impact of uncertainty on sourcing performance paves the way to managing performance uncertainty.

But what does visibility to the impact of sourcing uncertainties on future sourcing performance mean, and how can it be achieved? It begins by acknowledging the presence of uncertainty in the sourcing environment and capturing that uncertainty through forecast scenarios. The fact that forecasts may "always be wrong" does not make them any less necessary or even useful. It does, however, make it essential to acknowledge anticipated error. This error can be reflected through forecast scenarios (high, base, low) scaled to the level of error commonly experienced in forecasts or can include specific scenarios such as the introduction of a disruptive technology on the demand side or the collapse (financial or physical) of a strategic supplier on the supply side.

Large amounts of information about the uncertainty of future supply and demand are available today within companies. This includes both the "forward-looking" information available to sales, marketing, and procurement through interactions with customers and suppliers, and in the form of market research. It also includes historical information about how forecasts have differed from realized values in the past, how actual prices and supply variability have varied over time, and how large scale events have affected both demand and supply. It also includes access to increasingly available and sophisticated databases and services that track natural disasters and other disruptive events around the world.

Forecasting teams are generally receptive to the range forecasting approach for two reasons: First, they are often frustrated by current forecasting processes that require large amounts of time and effort to generate a "best guess" forecast, which they are confident will be wrong. Moreover this "point" estimate forces them to exclude information about other possible outcomes that clearly have business value. Second, the direct link between "best guess" forecasts and many important resource allocation decisions and performance measures creates political influences that can distort the forecast and polarize the forecasting process. Range forecasts address both concerns by allowing information about all potential forecast outcomes to be captured and utilized, and generating range forecasts eliminates the link between forecast values and business decisions.

In summary, creating a formal process that acknowledges the risk, enables cross functional discussions about the nature of the risk, and ultimately quantifies the risk are necessary fundamental building blocks to building a risk management process. Once these building blocks are in place, the organization can then turn to designing supply chain strategies that mitigate these risks.

19.3.3 CAPABILITY #3: QUANTIFYING SUPPLY CHAIN PERFORMANCE AND DESIGNING SUPPLY STRATEGIES ACROSS UNCERTAIN BUSINESS OUTCOMES

As used here, the term "sourcing strategy" refers to the set of sourcing contracts or production decisions a firm may have (or is considering) for the material, and refers to the way the firm plans to use those contracts or production assets under alternative business scenarios. The objectives of these sourcing strategies can include everything from flexibility provisions that address the everyday uncertainty to disaster recovery provisions. The development of this capability can be broken into three sub-capabilities: (1) quantifying sourcing trade-offs, (2) analyzing supply chain implications, and (3) structuring and managing the sourcing relationships.

19.3.3.1 Quantifying Sourcing Trade-offs. Referred to as the quantify capability hereafter, this ability is central to systematically managing and monitoring supply chain risk. With detailed data, management can clearly assess the performance implications of prospective sourcing strategies over a range of outcomes and choose those that match risk/reward profiles. This assessment serves two important business purposes. First, it enables currently proposed strategies to be refined using information gleaned about performance trade-offs and the specific business impact of alternative contract structures. Second, it enables efficient and effective sourcing management by providing quantitative measures on which to base decisions, internal communications, performance expectations, and performance measures.

Consider the case of a specialty chemical manufacturer who originally located one of their biggest plants close to their biggest source of supply by the Gulf of Mexico. Unfortunately, the plant loses significant capacity each year due to hurricanes. They are now exploring a series of alternatives, including substantial capacity and inventory investments to systematically position assets to mitigate this risk. All of these come at some additional cost and most executives need to be able to quantity the risk-reward trade-offs for buying what is effectively supply chain insurance.

The output generated by the SCRM "quantify" capability for each prospective supply and demand outcome is comparable to a standard performance planning process, and includes values for each of the key operating and financial performance measures over the time period analyzed. The difference under SCRM is that this information is generated for each potential supply and demand outcome, rather than simply for the unlikely scenario that everything goes according to the "best guess" plan. This makes it possible to analyze and specify performance objectives much more completely, including flexibility and risk, at levels that the data shows are challenging but achievable.

For example, objectives may include specified trade-offs for cost, availability and liability across supply and demand outcomes, and comparable objectives for cost, flexibility and risk for other key variables. Having specific values for metrics such as these available at the planning stage, coupled with the ability to review, negotiate, and modify them with key stakeholders as appropriate, ensures the hard issues can be recognized and resolved, and that clarity, alignment, and accountability are achieved.

Imagine a cross functional team from marketing, manufacturing, and procurement responsible for planning a critical new product launch. After reviewing SCRM performance data across the prospective supply and demand outcomes, the team determines that the company's business goals are best served by ensuring it will be able to meet demand at or near the highest projected demand outcome, should it occur. As a consequence, however, they acknowledge that clearly quantified levels of incremental cost and liability will result if the product underperforms. The team communicates both plan and performance data for representative high and low demand scenarios to finance and general management for review and approval.

However, after following a similar process, the same group may arrive at very different performance objectives for an established high volume, low margin product. Appropriate objectives are likely to focus on minimizing cost and margin risk across potential outcomes, and may specify caps on inventory and other liability risks at levels determined using the SCRM data. In contrast to cost and liability risk, the key trade-off to communicate to management for review and approval may be quantified product availability risks under high demand scenarios.

19.3.3.2 Analyzing Supply Chain Implications. The supply chain architect function supports an SCRM program with two primary sets of deliverables. The first is to identify the points in the supply chain that present the most risk. These points of exposure may be the result of the technology involved, the location of the supply, the mode of transportation, the manufacturing process, or the supplier. The second is to design the set of options that are available to address the traditional cost, availability, and inventory metrics and that will address the risk management objectives. The range of options that may be considered are investments in extra capacity, qualifying additional product facilities, holding reserves of strategic materials at multiple supply hubs, or changing product design. In doing so, this function can capture the cost and performance attributes of different options, which is the necessary data to quantify the sourcing trade-offs.

It also is important to consider how suppliers are affected when their customers begin to tailor their sourcing strategies to more precisely reflect their sourcing objectives and risk management requirements. The direct impact of this change is that suppliers receive much more specific information about the nature of the sourcing performance required, and the value of such structured "supply services" to both the supplier and the organization it serves. Many suppliers may find that these changes significantly alter their competitive landscape and that they may have to invest in a new and significant set of assets or capabilities. For example, any supply base that is "over-exposed" to a particular region based on labor costs, may now find that their lack of geographical diversity may make them less attractive. At a more tactical level, suppliers will find themselves having to respond to requests for detailed recovery plans and details such as time required to resume to full production rates after a catastrophic event or a significant labor strike. As it stands today, large portions of the supply base have little or no capabilities to respond to these types of requests and most companies have no way of validating these responses.

Alternatively some suppliers will see this is as an opportunity. The risk management framework gives them an opportunity to compete on additional dimensions beyond cost and functionality. Suppliers that have developed the core risk management capabilities can begin to capitalize on these investments more directly and quantify the value they are providing to their customers. Anything that enables them to compete on dimensions other than price almost always leads to improved margins.

19.3.3.3 Structuring and Managing Supply Chain Relationships. Once performance objectives and trade-offs have been set across the range of potential supply and demand outcomes, the next SCRM process step is to determine the supply chain resources–the assets, strategies and relationships—that best achieve them. This is achieved by quantifying the performance for each set of supply chain resources under consideration, and identifying the one that performs best.

Perhaps the most important benefit this quantification capability provides is that it allows the organization to determine the best possible supply chain alternatives and the supply chain relationships that define them. Supply chain staff members are free to assess the costs and capabilities of prospective supply chain partners and initiatives, and identify and negotiate the specific performance terms that best meet requirements—in short, to provide the high-value services they'd like to be able to spend their time on, and which define the potential of the supply chain.

By drawing on the information generated through this quantification capability, these teams can complete these activities much more efficiently and effectively. In the same way the SCRM data allowed internal teams to set and gain alignment around specific performance objectives and trade-offs across potential supply and demand outcomes, it also allows companies to be clear and specific with their supply chain partners. However, each external partner typically only comprise one element of the overall supply chain. Supply chain and procurement teams must evaluate the external partner as one element of the overall supply chain and determine its impact and interactions with other elements/partners of the supply chain.

A second component of the data generated by the quantification capability quantifies how individual supply chain resources are utilized across supply and demand outcomes. This includes both where they do and do not meet requirements, and where they align and fail to align with the other resources secured. For example, this data may show that flexibility or disaster recovery plans are not aligned across a critical set of materials or manufacturing capacity. If one part of the sourcing team has secured more flexibility or resilience in their supply agreements than other portions of the sourcing team, then the company is effectively paying for insurance that will not provide any value. These imbalances are more likely to occur in nascent efforts to establish disaster recovery plans with a global supply base where companies typically have not yet invested in the tools and processes to systematically track and monitor their disaster recovery plans.

As a result of this process, suppliers—both as a group and as individual components that rely on the performance capabilities of other partner companies—benefit from commitment to the more specific and mutually aligned terms that are generated. Individually, suppliers are able to use the more specific information and the refined incentives they receive to more effectively plan and put appropriate resources in place in advance. This, in turn, enables them to better balance cost, availability, lead time, and liability over time in response to actual supply conditions. This results in reduced costs and improved performance. As a collection, each supplier is more insulated from seemingly unrelated disasters that suddenly affect their business. To understand this, you only need to consider the fate of a supplier in one region whose orders disappear because their customer's lines have been shut down by a supplier in another region who has been shut down by an earthquake.

Having laid out a set of capabilities that will support a supply chain risk management approach, it is now important to see how these capabilities can be integrated into a repeatable and scalable process.

19.4 Process Approach to Supply Chain Risk Management

In 2006, representatives of several organizations came together to form the Supply Chain Risk Leadership Council (SCRLC). The SCRLC's ultimate goal is to establish mutually agreed-upon, industry best practices, process frameworks, and tools to aid in the development of effective supply chain risk management. The Council includes leading companies around the globe, represents a broad cross-section of industries, and is recognized as the premiere thought leader for supply chain risk management.

Through member collaboration, the SCRLC has developed an SCRM system and process approach. The approach is similar to the "Plan Do Check Act", or PDCA model, and involves six distinct process steps. As depicted in the diagram below (Figure 19.2), each of the SCRM process steps continuously gain new

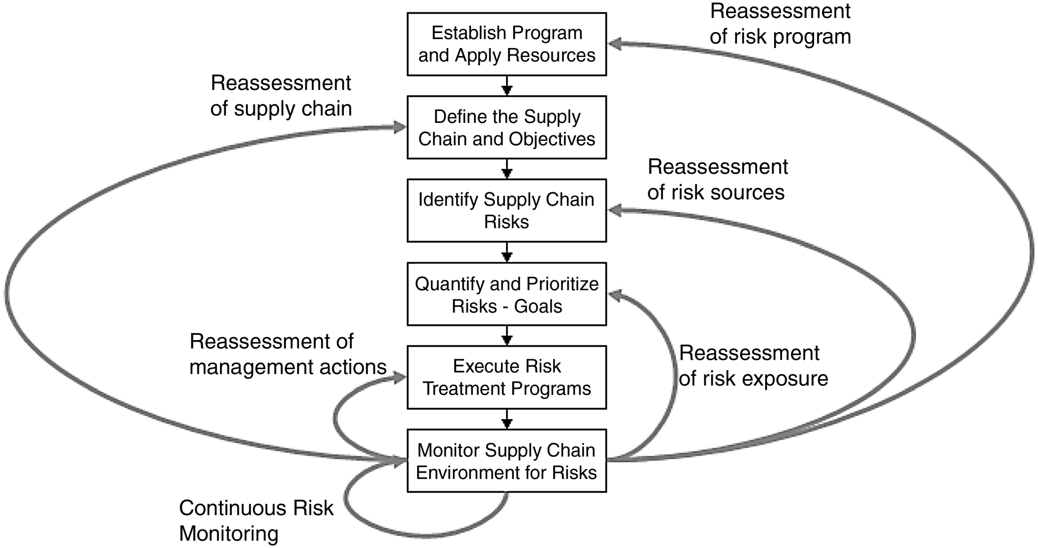


FIGURE 19.2 TheSCRM Process Approach *Source:* Supply Chain Risk Leadership Council, 2009.

information by incorporating learning from the global environment, ensuring continuous improvement and effectiveness.

19.4.1 ESTABLISH PROGRAM AND APPLY RESOURCES

As the first step in the SCRM process, the organization must recognize risk management as a priority. To do this effectively, an organization must secure executive support for risk its management program. Once executive support is established, the organization must scope and secure resources necessary to execute its risk management program objectives.

A big component of establishing an SCRM program is implementing the right set of performance measures. There are distinct aspects to measuring an effective SCRM function. First, the measures have to focus as much on good decisions as they do good outcomes. While this may be universally true, in practice most measures only focus on good outcomes. However, in a world in which there are multiple planning scenarios, the right decision to hedge against uncertainty will also affect bottom line performance if nothing actually happens (no different than regretting buying auto insurance if you have never had an accident). Second, in an effort to create internal risk management plans and to solicit risk management plans from suppliers, it is important to have a set of measures that captures their performance to these stated plans. In the past, without these plans all parties simply claimed they were doing the best they could and there was no means to enforce accountability or to assure consistency in the quality of response across multiple suppliers.

19.4.2 DEFINE THE SUPPLY CHAIN AND OBJECTIVES

As stated previously, understanding the supply chain environment is a critical activity, and is often done to some degree before a program is realized. It is this understanding of the supply chain and the risk exposure that will likely drive senior management to establish a program. Other drivers are of course an unexpected crisis event that grabs the attention of corporate officers, board members, investors, or the customer. The objectives of the organization and that of its internal and external suppliers and customers are essential to understanding the motivation that drives behavior.

19.4.3 IDENTIFY SUPPLY CHAIN RISKS

The process of identifying supply chain risks within the organization and that of its global environment is essential to conducting a comprehensive risk assessment. Risks will vary depending on industry, and may be many tiers away from an organization's internal responsibilities. An example would be the identification of a Tier 4 supplier of raw material being mined in a remote part of the world, for which your product contains sole sourced components that rely on its uninterrupted supply.

This part of the process is closely tied to the second and third core capabilities. The second capability focuses on the identification of the sources of the uncertainty or risk. Within the third core capability, the "analyze supply chain implications" activity (see Section 19.3.3.2) comprises identifying the points of exposure to these risks within the supply chain and formulating initial strategies to mitigate these risks.

19.4.4 QUANTIFY AND PRIORITIZE RISKS AND GOALS

After establishing a program, mapping the supply chain and identifying risks, the task of risk quantification will undoubtedly be the most difficult to accomplish, due largely to the availability of trusted risk probability data. Many organizations struggle to move beyond the quantification step because they cannot reach agreement, which is essential in prioritizing risks and taking the step toward moving into a more proactive role of applying risk treatments. In the supply chain, special emphasis should be placed on the time to recover from a disruption, which can be translated into lost revenue very easily. The probability of disruption types and the likely duration of the event will require data analysis, benchmarking, and sound judgment. Overall, an organization must establish a repeatable process to determine the appropriate measures and weights in determining relative risk and priority.

This part of the process also is largely supported by the second and third capabilities. Specifically, the quantification of risk and goals part of this process mirrors capability 2 (see Section 19.3.2), which focuses on a framework for quantifying the sources of uncertainty and communicating this information in the form of scenarios and/or range forecasts across the organization. Similarly, prioritize goals and objectives portion of this process is closely linked to the "structuring and managing supply chain relationships" sub-capability (see Section 19.3.3.3). This portion of the process and the capability focuses on how to coordinate risk management strategies across the organization in a manner such that they meet the organization's priorities and goals.

19.4.5 EXECUTE RISK TREATMENT PROGRAMS

Risk treatments can take many forms and be applied anywhere within an organizations global environment. Treatments will be funded and applied as management sees appropriate, which will be largely based on the organization's willingness to accept risk, or risk appetite. Additionally, if the organization has not effectively quantified and prioritized its risks, there will be reluctance to invest in expensive risk treatment programs. Some examples of risk treatments include qualifying an alternate manufacturing site, added security in a key facility, securing a second source for a long-lead component, or investing in a small business that has become a strategic supplier for one of your new products. Most organizations will not be able to treat all risks, so it is important to choose a strategy and make continual effort toward achieving the particular goals established for your organization.

19.4.6 MONITOR THE SUPPLY CHAIN ENVIRONMENT FOR RISK

This last step in the process is largely about the first core capability: collect, communicate, and respond. By continuously monitoring (i.e., collecting) the supply chain for potentially disruptive incidents, the organization can significantly reduce the time required to react to a real crisis. This collection, combined with a defined response playbook (communicate and respond), provides the necessary pieces to immediately and effectively execute the risk management program.

Continuous risk monitoring ensures the SCRM process is able to adapt and change with the ever changing business environment. Other changes to consider include the introduction of new products, suppliers, partners, transportation lanes, customers, etc. It is not uncommon for an organization to adjust its focus based on real events. A large hurricane for example may highlight a previously unknown or under-weighted risk at a key supplier site. Similarly, over time, factors such as a global recession may impose a strain on one or more of your supplier's ability to secure credit, which would place a larger emphasis on financial health of your suppliers as part of the overall risk program and investments required to drive the necessary risk treatments.

19.5 Case Study: Cisco Responds to the Sichuan Earthquake

19.5.1 RISK MANAGEMENT GROWS IN IMPORTANCE AT CISCO

Over the last few years, the SCRM team has become a strategic element of Cisco's Customer Value Chain Management (CVCM) organization. CVCM is a central function that collaborates with other Cisco teams and external partners to plan, design, manufacture, deliver, and ensure the quality of the company's products and solutions. Formerly called Global Supply Chain Management, CVCM acquired its new name recently as part of a broader reorganization to better focus on the total customer experience.

In order to drive this program, the SCRM team partners closely with several other CVCM functions, including Global Supplier Management (GSM), which oversees sourcing decisions and manages relationships with Cisco's component suppliers globally; Product Operations, the function responsible for transforming engineering innovation into robust products; and Global Manufacturing Operations, which oversees the company's global manufacturing and logistics operations through a network of outsourcing partners.

19.5.2 CISCO'S SCRM TEAM FOUR KEY PROGRAMS

1. **Business Continuity Planning (BCP)** The BCP program is the foundational element of the Cisco SCRM management system. The BCP program collects and maintains critical information from Cisco's key suppliers,

contract manufacturing partners, and transportation and logistics providers that is used to define the supply chain and identify risks. The collected data include information such as supplier site locations, single and sole sourced component manufacturing sites, site recovery plans, and recovery lead times. These data drive a set of resiliency measurements by supplier, partner, and product, which are used to drive continuous improvement. These measures are used to quantify and prioritize risks and goals for the organization.

2. **Product Resiliency** The Product Resiliency program leverages the component BCP data to drive product-based risk treatment programs. This program helps Cisco's business units make informed and strategic sourcing decisions throughout the product lifecycle, by proactively mitigating component risks, such as qualifying alternative sources for long lead time sole sourced components, and qualifying alternate supplier sites. One of the risk mitigation strategies of the Product Resiliency program is to focus on single and sole sourced components that touch a large percentage of Cisco revenue and have a time to recover that is outside of the program's stated resiliency objectives. Besides revenue, other factors for product risk mitigation include high profile customers and other factors such as supplier financial stability.
3. **Supply Chain Resiliency** The Supply Chain Resiliency program uses the BCP data to identify nodes in the supply chain with recovery times outside of Cisco's risk tolerance. This team partners with Electronic Manufacturing Services (EMS) partners and transportation and logistics suppliers to develop resiliency plans and will identify and drive risk treatment programs to improve the recovery times for long-term and short-term disruptions. Example risk mitigation tactics include tropical cyclone preparedness checklists, backup power capabilities, and alternate site and route options for manufacturing and logistics respectively.
4. **Supply Chain Monitoring and Crisis Management** Cisco's global crisis management team is responsible for monitoring the global supply chain and responding to potential and real disruptions on a 24/7 basis. One of the key elements of this program is the supplier site mapping capability, which is made possible by the BCP program. By joining supplier site, component data, and event locations, news around the globe—ranging from hurricanes to political unrest—can be filtered and monitored with little effort. The result is a management system that is able to respond in almost real time to events that have possible impact to Cisco's supply chain. In addition to the monitoring capability, the Crisis Program has built a robust crisis management team, consisting of functional experts across the organization that are trained to respond and manage supply chain disruptions.

19.5.3 KEY BUSINESS PARTNERS

Cisco's SCRM program requires a truly collaborative effort to deliver resiliency for Cisco's highly complex supply chain. To illustrate, SCRM partners with Cisco engineering to assess the resiliency attributes for new products. This engagement occurs well in advance of "first customer ship," giving development engineers time, if needed, to consider alternate or more resilient components before the design is finalized.

Similarly, SCRM engages with the product operations and manufacturing operations functions to assess the resiliency attributes of the anticipated build-to-ship supply chain. This forward-looking assessment allows Cisco to incorporate supply chain resiliency as a consideration in supply chain design and business awards to EMS partners.

For sustaining products, the team has developed a highly collaborative model with GSM and manufacturing operations, working closely to define the resiliency programs that need to be executed. However, once these programs are scoped, it falls upon the GSM and manufacturing operations teams to do the heavy lifting by working with Cisco's component suppliers and EMS partners to implement the applicable resiliency program (for example, qualifying second sourcing and alternate sites and negotiating and implementing buffers).

In certain cases, design mandates are at odds with optimal supply chain resiliency. There can be some inherent risks in designing in a product from a start-up supplier—a company that has no demonstrated ability to ramp, deploy a robust BCP, troubleshoot quality or maintain financial stability over time but provides required technology to differentiate product functionality. By designing a plan to mitigate against these elevated risks, SCRM enables the company to move forward, to adopt an innovative component it might not otherwise have had the risk tolerance to use.

19.5.4 RISK MANAGEMENT PROGRAM IN ACTION

Cisco has hundreds of suppliers producing components at thousands of sites that feed its EMS partners around the world. With 95% of its production outsourced, Cisco's supply chain footprint is very global. To enable supply chain monitoring, the team first had to understand the footprint, or more specifically, where Cisco components were being built.

To collect this critical information, the SCRM team developed the BCP program, which collects key information required to perform a supply chain risk assessment, in addition to an effective crisis response. Data collected from suppliers include: physical address of supplier sites, emergency contacts, alternate manufacturing locations, and time-to-recover to an alternate site. The continuity planning process also includes gathering data to evaluate a supplier site's own business continuity plans, or readiness and resiliency in the event of a supply chain disruption.

With the footprint defined, the SCRM organization was then ready to begin correlating world events to strategic locations on the map. The team utilizes NC4 (National Center for Crisis and Continuity Coordination), which allows it to build alert profiles based on specific locations or geographies. NC4 customers, such as Cisco, are then able to subscribe to alerts based on a set of filters for attributes including event severity and event type.

The team's BCP program and monitoring plans were put to the test on a sunny afternoon on May 12, 2008, when China was devastated by one of the largest and most destructive earthquakes in modern memory. With a magnitude of 7.9, the quake's epicenter was located a mere 80 km northwest of Chengdu, capital of Sichuan province. Final estimates indicate as many as 68,000 were killed as a direct result of the quake, and an additional 4.8 million were left homeless.

In the case of the Sichuan earthquake, the team was alerted based on the following NC4 profile: *Moderate to extreme meteorological and geophysical incidents within 100 miles of a supply chain location.*

This real-time monitoring, coupled with supply chain locations, provides the SCRM team near-immediate notification of incidents and greatly shortens the response time to events that are out of the team's control. In most cases, many companies are learning about impacts to their supply chains days or even weeks after incidents such as earthquakes and labor disputes occur.

For Cisco, the Sichuan earthquake presented a potentially high-stakes test for our new supply chain risk management framework:

Within 48 hours of the earthquake, Cisco was able to conduct a full impact analysis, including evaluating affected supplier sites, parts, and products. The robust BCP platform allowed the team to gain complete visibility into the supplier footprint in the area. Within two days of the earthquake, SCRM had initiated a crisis survey targeted at the suppliers' emergency contacts in the region. Meanwhile, the crisis team had partnered with affected Cisco organizations and reviewed any potential revenue impact.

The analysis performed within the first 24 hours revealed that Cisco had approximately 20 suppliers in the affected area. While there was no impact to any of the manufacturing sites and logistics centers, there were two suppliers potentially at risk: Supplier X, which presented a significant revenue exposure for Cisco in addition to the risk of being single sourced, and Supplier Y, with a smaller revenue impact but with physical damage to one of its buildings.

Coincidentally, the SCRM team, in conjunction with the GSM function, had already been proactively working to address the single-source risk with Supplier X and had identified a second source a few months prior to the earthquake. However, the situation with Supplier Y remained an issue.

The Crisis Management team engaged its internal sourcing, planning, and operations colleagues to deploy previously identified second sourcing options, as well as to gain commitments from the supplier for additional capacity.

Despite facing a natural disaster of huge proportions, Cisco was able to respond rapidly, ensure the safety of the extended supply chain, identify the risk exposure to the company, and work with its EMS partners to mitigate the risk, thus ensuring no impact to customer shipments. The continuity planning, crisis management, and risk mitigation arms of the SCRM team worked in close collaboration with internal partners in this endeavor.

19.5.5 QUANTIFYING THE IMPACT

Once the sites affected by the earthquake in China were identified, the team could quickly leverage and combine the BCP information with supply chain visibility data to determine any potential impact to Cisco's customer shipments or financial bottom line by quantifying the exposure. The BCP data provides time-to-recovery data for each of the supply chain locations, including raw material suppliers, logistics and transportation providers, and EMS partners. This enables visibility into which components, materials, or products are produced at each supplier locations. In addition, the supply chain data allows the Cisco SCRM team to determine which products and how much revenue is enabled by each of the logistics and EMS partner locations.

The Cisco team was able to leverage these same analytical capabilities to develop supply chain risk assessments, helping the company focus on proactive risk mitigation programs with the right priorities.

To accomplish this, Cisco uses a "risk engine" to assess the likelihood of a disruption. The risk engine incorporates many data sets (such as 100 year flood data, actuary data, geological and geo-political data, site incident data, and supplier performance data) to assess the likelihood of a disruption. These disruptions are correlated to Cisco supply chain locations including supplier sites, contract manufacturing facilities, and logistics centers. The impact of a disruption is determined based on the revenue enabled by each node in the supply chain and that node's recovery time. Cisco also uses simulation capabilities to integrate all of these data sets into a single model that generates "heat maps" based on likelihood and impact.

19.5.6 MITIGATING RISK IN THE SUPPLY BASE

Supplier and component resiliency has always been core to Cisco; for this reason, the company takes a very proactive approach to ensuring that whenever possible its products have two or more sources qualified for each part. GSM and Product Operations have dedicated resources and funding to identify and "de-risk" single sourced and other risky components. Each part has a risk attribute that identifies its sourcing status (single-sourced vs. multisourced), quality history, technology status (legacy vs. new) and lifecycle (new vs. end of life). These risk and other, component-level attributes provide design guidance for new products. A key function of component teams is to find and qualify alternate sources for the single-sourced parts. Where second sources cannot be found, Cisco has dedicated resources and funding to develop an alternate source for certain key components with very high impact to revenue.

Cisco launched the Global Component Risk Management (GCRM) system to formally streamline the efforts of the diverse groups engaging in risk mitigation efforts throughout the organization. The management system prioritizes their efforts and provides an automated tool to log history of risk mitigation efforts at the component level. It also provides the capability to track and manage progress and status of risk mitigation by different groups. This approach eliminates duplication of efforts, establishes clear ownership and target completion dates, and helps keep track of mitigation activities being pursued on thousands of components in one central repository.

19.5.7 MITIGATING RISK IN MANUFACTURING FOOTPRINT AFTER THE SICHUAN EARTHQUAKE

Based on the impact assessment, Cisco was able to quickly determine if there was any impact on manufacturing, transportation, and/or logistics nodes resulting from the earthquake. If there had been disruptions, the company's proactive risk mitigation program would have identified alternatives and ensured that they could have been enabled quickly through product qualifications and defined recovery plans that achieve the stated recovery time objectives.

In order to achieve mitigation, the SCRM team assesses the current recovery capabilities and identifies gaps that could limit Cisco from recovering within the desired timeframe. These gaps form the basis of the mitigation programs. Specifically, Cisco works with its EMS partners to close these gaps through developing work-around processes, reducing equipment lead times, and enabling quick ramp-up at an alternate facility.

For example, to mitigate the risk around the EMS partner locations, Cisco looks at developing recovery plans with their partners, including agreements for additional capacity that we can leverage in the event of a disruption. The goal is to integrate risk requirements into the company's capacity planning processes. For test equipment with a lead time longer than the recovery objective, the SCRM team will work with a supplier to determine the appropriate mitigation solution. These solutions can range from setting lead time agreements to purchasing inventory of long lead time materials and securing burst capacity to meet demand surges.

For logistics centers, the team works with the EMS partners to identify additional space and/or facilities that can be leveraged, including redundant warehouse processing equipment. These proactive mitigation solutions and recovery plans became useful during the Sichuan earthquake. They allowed Cisco to leverage alternate transportation solutions, offer additional transportation capacity, and expedite capabilities to the impacted suppliers in the region.

19.5.8 THE SLIVER LINING TO DISRUPTIONS

Disruptions provide a unique opportunity to enhance your capabilities. While the core mission of any supply chain risk management program is to mitigate disruptions (if not avoid them altogether), there is a silver lining to even the most unfortunate occurrences. Process development, mitigation programs, and even drills will only tell you so much about your supply chain's readiness and responsiveness in the case of a disruption. The true test of supply chain resiliency comes in the face of events like the Sichuan earthquake. Even the most thoughtful and thorough crisis management program will not anticipate every aspect of a disruption. Moreover, disruptions tend to be idiosyncratic in nature—each taking on dimensions and requiring a different response tailored to that event. In the face of real disruptions, however, organizations are presented with unique opportunities to refine their programs.

In the case of the Sichuan earthquake, Cisco learned several things about its own program and made appropriate changes. As a result of the event, Cisco revised the membership of its crisis management teams to include key external players like manufacturing and logistics partners and adjusted the activation process. The company also made its risk management playbooks and trigger points more forward looking and proactive based on the type of event. More importantly perhaps, the team gained a fuller appreciation of the importance of having a closed-loop post-mortem process that allows them to capture key lessons and evolve the program. Cisco understands that a program will never be "baked," but rather must be capable of incorporating useful learning from new experiences.

19.5.9 SECURING SENIOR MANAGEMENT SUPPORT

A SCRM program cannot function in a vacuum; it needs to be a top priority at the highest level and across the organization. Despite the overwhelming presence of persistent risk in supply chains, however, only 12% of companies report having a risk-resilient global supply chain.[3]

Support for Cisco's program comes from the very top. CEO John Chambers and Angel Mendez, senior vice president of the CVCM organization, are staunch supporters and maintain an active role in promoting and driving ongoing attention to supply chain resiliency. Chambers is briefed quarterly and sees SCRM as an integral part of the corporation's risk profile. Mendez, too, has great passion around SCRM and has often found himself in areas where a crisis is going to develop. For example, he was in Hong Kong in the midst of severe typhoons last year, and when the H1N1 flu virus broke he was in Mexico City, as part of a Cisco delegation meeting with the Mexican president.

For Cisco, events like 9/11 and Hurricane Katrina were catalysts. Even though these events did not significantly impact Cisco, senior management saw the potential for disruption resulting from supply chain issues. Thus, the SCRM team was formed to proactively put in place infrastructure, processes, programs and tools to prevent, thwart, and recover quickly from a major disruption.

Building resiliency can be an expensive proposition. After all, it's a highly complex undertaking. In Cisco's case it involves collecting BCP data from over 700 suppliers, identifying and qualifying second sources on thousands of single-sourced components, and building a crisis response capability that enables a coordinated response for more than 8,000 products. Dedicated funding is required, with a budget set aside each year with a clear set of objectives to be accomplished.

Finally, the team has developed metrics that enable Cisco to measure resiliency at the product, site, regional, geography, and business unit level. In particular, Cisco developed a "Resiliency Index," which is a composite of resiliency attributes that is calculated and reported by the business units at Cisco's semiannual operations review with senior management. By having a resiliency metric that is shared among the business units and supply chain, CVCM have driven awareness and understanding of what resiliency means to Cisco as well as a common framework for driving improvements.

---

[3] Aberdeen Group Report, *Supply Chain Risk Management: Building a Resilient Global Supply Chain*, July 2008.

19.6 The Importance of an International Standard in SCRM

Currently in the global business community, no single, international standard exists that defines the steps, processes and goals an enterprise must undertake to ensure the establishment of an effective SCRM system. As a result, there is no comprehensive structure an organization can use to evaluate the resiliency of a partner company, or a supplier, and thus, no substantive way to ensure true supply chain stability and resilience across organizational and international boundaries. An International standard would enable interested companies to:

1. Establish, implement, maintain, and improve a resilience management system.
2. Assure itself of its conformity with its stated resilience management policy.
3. Demonstrate conformity to its customers by seeking certification/registration of its resilience management system by an external organization.

Many experts today believe the creation of an international standard, as certified by the ISO (International Organization for Standards) is necessary to truly drive resiliency. An ISO SCRM standard would serve to level the playing field, ensuring that companies with global operations are adhering to one set of guidelines across their operations, and provide companies who want to develop a resilience approach with the basic components: how to organize activities, processes, functions, and management systems, as well as benchmarking that will aid in determining how to measure success.

The adoption of such a standard would take many years, but would likely have the penetrating and effective impact that ISO9000 has done for quality and ISO14000 has done for the environment. Today the 9000 and 14000 standards signify a level of discipline and responsibility that sets suppliers apart.

19.7 Conclusion

Companies like Cisco have fully invested in supply chain risk management and for their suppliers it has become a necessary ingredient of doing business with them. Building an effective SCRM program begins with top management support, especially if a company's customers are not mandating a level of resiliency prior to awarding new business. Once a firm has committed to building a more resilient supply chain, it needs to create the ability to collect, communicate, and respond to new information; assess the impact of uncertainty; and quantify supply chain performance across uncertain business outcomes. Achieving this requires a management system and commitment from the organization to make resiliency a top priority along with safety, quality, and social responsibility. Today's global supply chains are exposed to new risks, which leave not only our own companies at risk, but that also risk the global economy. The focus on cost reduction and quality has resulted in entire regions that specialize in certain types of technology and manufacturing, such as Taiwan for semiconductors. Companies are now beginning to acknowledge these risks and are investing in SCRM programs, but risk treatments come with a cost. Alternate components often require additional certifications, inventory carrying costs are constantly under pressure to be reduced, while inventory turns are under pressure to improve, and manufacturing capacity is constantly being adjusted to meet demand. Companies need a reason to invest in building resiliency into their supply chains. If it is not a requirement of doing business, unacceptable levels of risk will exist somewhere.

CHAPTER TWENTY

A Bayesian Framework for Supply Chain Risk Management Using Business Process Standards

CHANGHE YUAN, FENG CHENG, HENRY DAO, MARKUS ETTL, GRACE LIN, AND KARTHIK SOURIRAJAN

Supply chain risk management (SCRM) has gained significant attention in recent years. While much of the existing literature focuses on proactive aspects of SCRM such as risk identification, assessment, and mitigation, little work has been done regarding enabling quick actions after risk events. The challenge lies in the lack of a SCRM framework that would support the categorization and integration of complex risk factors, heterogeneous information, and hierarchies of business

Acknowledgments: The first author, Changhe Yuan, is supported by National Science Foundation grants IIS-0953723 and EPS-0903787. The modeling and analysis applications for the case study were developed based on the proposed Bayesian Model and GeNIe/SMILE, a probabilistic graphical modeling tool developed at the Decision Systems Laboratory and available at http://genie.sis.pitt.edu.

Handbook of Integrated Risk Management in Global Supply Chains, First Edition. Edited by Panos Kouvelis, Lingxiu Dong, Onur Boyabatli, and Rong Li.

processes and their interrelationship. In this chapter, we describe a framework and methodology for supply chain risk analysis that uses Bayesian graphical models to identify, quantify, mitigate, and respond to the risks affecting a company's global supply chain. The proposed methodology uses a *two-dimensional network categorization of information* about risk factors and business processes. It allows managers to effectively specify the risk environment by mapping all risk variables to business processes. The framework is designed to automatically learn the underlying risk models in a well-structured fashion using historical supply chain data to obtain qualitative and quantitative interdependencies among risk variables. The resulting risk models allow analysts to identify high risk areas in a supply chain business process, diagnose risk factors contributing to observations regarding abnormal events, and analyze the sensitivity of supply chain performance measures to various risk factors. The models provide guidance for identifying risk mitigation strategies and for responding to disruptive events. We illustrate the methodology using a comprehensive case study based on global logistics process performance data.

20.1 Introduction

Even before the recent global economic downturn, supply chains were on an arc of continued growth in complexity. Corporate executives faced extended and intricate supply chain processes and challenging environments and had to reach into new territories to search for opportunities for cost reductions and agility. The economic downturn and the resulting volatility have exposed companies to amplified supply chain risks seldom seen before, far broader in scope and greater in potential impact than in the recent past.

Many events can affect supply chain operations. These can be *environmental* such as natural or man-made disasters or vary according to socio-political conditions; *intra–organizational,* such as labor disputes, plant accidents, bankruptcy procedures, or operational concerns; *inter-organizational,* with concerns regarding communication, collaboration, and coordination among partners, suppliers, customers or other companies. The impact itself can be *short-term* and related to cost, supply, demand, and quality; or *long-term* and related to business continuity, growth, or stock prices.

There is a definite heightened perception of supply chain risk within the industry as risk is visibly on the rise. An increasing number of incidents have illustrated how local events can quickly result in significant global impact. Table 20.1 shows a few such incidents. An Aberdeen Group research study (Aberdeen 2008) indicates that 88% of executives reported fragile supply chains, and 46%reported the need for better SCRM. They also reported that 77% of executives noted increased concerns about supply chain(s) versus 65% in a 2006 report.

A recent IBM study (IBM 2009) indicates that the top two concerns of supply executives worldwide are *supply chain visibility* and *managing risks,* both operational and financial, to reduce the likelihood of supply chain failures. Indeed, the present environment makes it crucial to gain near real-time visibility and to

TABLE 20.1 Recent Examples of Supply Chain Failures

Company	Supply Chain Risk Area	Impact
Ericsson	Single source of supply for critical manufacturing element.	• Loss of $400 million in sales after fire at Ericsson's microchip supplier plant (Norrman and Jansson 2004).
Boeing	Complex global supply chain requiring many components to come together for successful production.	• Year plus delay in the delivery of the 787. • Cut in revenue forecast for 2008 of $500 million (Rigby 2008).
Toyota	Single source of supply for critical manufacturing element.	• Delay of 55,000 vehicles after earthquake shut down production at Toyota's piston-ring and transmission-seal supplier plant (Chozick 2007).
Sony	Manufacturing issues at plant producing key system component.	• Three-months delay in European PS3 launch. • Reduction of forecasted units shipped from 4 million to 2 million in 2006. • Reduction in profit targets from ¥130 billion to ¥80 billion (in Japanese Yen) for 2006 (Associated Press 2007).
Cisco	Ability to flexibly ramp up and ramp down production of time and technology sensitive components.	• Write-off of $2.5 billion in obsolete inventory stemming from purchasing components months before to production needs. • Excess inventory totaling $1.6 billion (Pender 2001).

identify and effectively manage risks that range from tremendous financial stress of suppliers to high-demand volatility and uncertainty, and increased competition causing pricing pressure. Many companies have initiated and implemented risk and performance monitoring, but executives are aware of the challenges ahead, which include the lack of integrated processes, data, enabling technology, and business analytics capabilities. Consulting companies such as IBM, Deloitte, Accenture, PWC, and PRTM see SCRM as a growth segment and are actively offering services to help their clients in that area.

Supply chain risk is defined as any uncertainty that can affect the organization negatively, such as financial loss or operational impact. The Supply Chain Operations Reference Model (SCOR), a process reference model developed by the Supply Chain Council (2008), is widely accepted as a cross-industry standard reference model for supply chain management (Supply Chain Council 2008). SCOR defines SCRM as the process of identifying, coordinating, and managing supply chain risks by aligning itself with an organization's overall business risk management program.

An effective SCRM strategy must address the following three major steps: risk identification and assessment, risk mitigation, and risk monitoring and control. These steps are summarized below.

Risk identification and assessment. The goal of risk assessment is to understand the risk environment of a supply chain business process, to identify and quantify relations among risk factors and their effect on the business value of the process, and to perform risk analysis to identify areas of high risk in business processes. The results can be utilized to develop and prioritize risk mitigation strategies.

Risk mitigation. The goal of risk mitigation is to reduce the impact of different root causes of risk factors with targeted strategies, some of which address:

- *Supply/demand* risk by redesigning the supply chain network for supply responsiveness, and by implementing the Sales and Operations Planning (S&OP) process to balance supply and demand;
- *Product risk* by managing product complexity and design for supply chain initiatives;
- *Demand risk* by customer rationalization, by managing revenue and price, and by improving demand forecast accuracy;
- *Supply risk* by sourcing contract management, manufacturing options planning, and hedging strategies;

Risk monitoring and control. The goal of risk monitoring and control is to continuously evaluate and improve SCRM strategies, mitigation options, and response. It consists of the following major tasks: monitoring key risk indicators, developing action plans to address the gaps on an ongoing basis, providing feedback on how well the risks are mitigated, and tracking value realization. Integrating risk management into supply chain planning and execution reduces vulnerability and builds a resilient supply chain. Companies can thus either prevent risks from occurring, or, at the very least, reduce the impact if they do occur, and shorten the recovery time needed. Risk mitigation and responses include first asking the following key questions:

- What are the supply chain risks based on the current supply to market strategy?
- What are the risks, returns, and potential response for each supply chain risk?
- What are the trade-offs between risk, response, and returns?
- What are the risk mitigation strategies?

Using the many existing sources of information, including expert knowledge, business process standards, and historical supply chain data, can help answer these questions. In the remainder of this chapter, we present a comprehensive end-to-end SCRM framework with a Bayesian approach used to support risk diagnosis, risk impact analysis, and risk mitigation and response. We also present a case study using the proposed framework and approach.

20.2 Related Literature

Supply chain risk management (SCRM) has recently received a great deal of attention from both industry and academia. As a result of the increasingly important role SCRM plays in today's global supply chain, related literature has grown significantly. SCRM is considered by many as a key component of SCM (supply chain management), which is not surprising given the uncertain nature of typical supply chains. SCRM covers a wide range of research topics and issues. It is not our intent to give a complete literature review but rather to provide a brief introduction to the literature closely related to the contents of this chapter.

Methodologies for supply chain risk assessment have been developed using statistics-based approaches. Kao et al. (2005) proposed to use dynamic Bayesian networks to represent the cause-and-effect relationships in an industrial supply chain. Some approaches (Deleris and Cope 2007, Treur 2008) rely on expert knowledge to construct a business process map and a Bayesian risk model. Alternatively, expert knowledge is used to understand the sources of risk and to construct the qualitative structure of the risk model with historical data that is subsequently used to estimate its parameters. One way of using expert knowledge to construct qualitative risk models is to collect risk-relevant information and incorporate it into traditional business process model diagrams, resulting in so called risk-extended business-process models (Deleris et al. 2008).

Research literature on supply chain risk mitigation is also extensive and overlaps with many research topics in SCM. Naim et al. (2002) describe a systematic approach for the collection and synthesis of qualitative and quantitative data from a supply chain in determining the vector of change. Zsidisin et al. (2004) look at supply chain risk mitigation from the perspective of a purchasing organization. They discuss supply chain risk mitigation techniques in terms of tackling issues arising from processes external to the organization, including strengthening supplier quality, lessening the possibility of supply disruptions, and improving the process through which vendors supply goods and services. Miller et al. (2005) describe a method for monitoring risks involving multiple document instances and risk categories. Feldman et al. (2007) describe an apparatus for identifying the components of a system most critical to the assembly of the final product and quantifying the impact on profit and revenue of the failure to effectively deliver these components. Cheng et al. (2007) have built a process flow model with deterministic links between risk variables for operational risk assessment and mitigation.

In the context of changing supply chain processes, Buchanan and Connor (2001) categorize supply chain risk in four areas: performance dips, project fights, process fumbles, and process failures. They further break down the process risk into a "people risk" category and an "operations" component. Hutchins (2003) argues that supply chain risk originates from areas external to the organization. He defines these risks as: supply chain partners' abilities to meet the contract, process and product requirements, the possibility of harm or loss if requirements are not achieved, the probability of an event with undesirable consequences, and the variation away from a specified set of requirements and how this variation is monitored and controlled. Based on the need for companies to adequately plan for business continuity, Finch (2004) looks at SCRM from the perspective of inter-organizational networking. He considered issues arising from processes both external and internal to the organization. Chopra and Sodhi (2004) categorized supply chain risk as arising from areas controlled internally by the organization including: manufacturing disruptions and delays, systems, forecast, intellectual property, procurement, receivables, inventory, and capacity. On a strategic basis, Christopher and Lee (2004) look at methods controlled internally including the need to increase supply chain confidence by improving end-to-end visibility across the supply chain as a mechanism for mitigating supply chain risk. An example of this is the sharing of demand forecasts in order to coordinate production and reduce the impact of demand amplification, referred to as the "bullwhip effect."

Tang (2006) proposes a framework called "supply chain risk and mitigation approach" for addressing SCRM issues along two dimensions. The first dimension addresses the risk level of certain events including both operational risks inherent to supply chains such as uncertain supply and uncertain cost and disruption risks caused by natural and man-made disasters such as earthquakes, floods, hurricanes, terrorist attacks, etc., or economic crises such as currency evaluation or strikes. The second dimension relates to the mitigation strategies that alter the impact of supply chain risks. Four basic approaches to supply chain risk mitigation that an organization can deploy through a coordinated/collaborative mechanism are discussed: supply management, demand management, product management, and information management. The study also provides a comprehensive review of various quantitative models for managing supply chain risks along with various SCRM strategies.

A recent study by Tang and Sodhi (2009) offers an updated view of current developments on SCRM. Their findings reveal and highlight three major gaps in the existing SCRM research literature:

- *Definition gap:* There is no clear consensus on the definition of SCRM. About 83% of the researchers surveyed believe that SCRM is a subset of SCM with no clear difference or with differences only in that SCRM focuses on risk elements.
- *Process gap:* There is a lack of emphasis on research on responding to risk incidents in contrast to their prevention and mitigation. The majority of the existing SCRM literature focuses on the proactive aspects of SCRM such as risk identification, assessment, and mitigation. Very few discussions relate to the reactive actions and strategies after the occurrence of risk events. Particularly, research on response to catastrophic events is rarely reported.
- *Methodology gap:* There is a shortage of empirical research in the area of SCRM.

We will address these gaps—particularly the process gap and methodology gap—in this chapter by exploring a two-dimensional SCRM framework, which utilizes the Bayesian approach and the SCOR model as the modeling and analysis methodologies.

20.3 A Framework for Supply Chain Risk Categorization

Risk management in a global supply chain is a challenging and complex task given that multiple sources of heterogeneous information including expert knowledge, business process standards, and historical supply chain operational data, need to be considered. Although advances have been made in recent years, major challenges remain. First, although experts typically have insights as to the major risk factors that affect a business process, eliciting expert knowledge is time-consuming and error-prone (Druzdzel and van der Gaag 2000). Second, operational data are often accumulated continuously over time for business processes and typically contain rich information for risk identification and quantification. However, the data are often heterogeneous, noisy, and incomplete. Finally, business process standards, such as SCOR (Supply Chain Council 2008), which provide cross-industry standards and excellent guidelines for supply chain management, only became available in recent years. All these sources of information are useful for risk assessment but no existing approach is capable of integrating all of them into an effective end-to-end SCRM framework.

To address the limitation, we propose a two-dimensional categorization framework with the dimensions being risk-variable categories and the standard business processes that allow an effective integration of these information sources in a well-structured way. In this approach, we only resort to experts in categorizing risk variables. Such knowledge is much easier to elicit from experts than qualitative and quantitative relations between the variables. Figure 20.1 is an example of the framework. We provide the detailed description of the two dimensions in the following sections.

20.3.1 RISK CATEGORIES

The first dimension of our categorization framework includes the three categories of risk variables—risk factors, risk events, and risk symptoms—plus performance measures. *Risk factors* are the root causes that may cause the occurrences of risk events to the business processes. *Risk events* are the disruptions, anomalies, or opportunities that may occur to a business process and have direct impact on the performance measures of the business processes. Some risk events are directly observable while others are not. Both observable and non-observable risk events can manifest themselves through *risk symptoms*.

The last category of variables is *performance measures*. Risk events have different likelihood of occurrence, but they also have different impact on the business

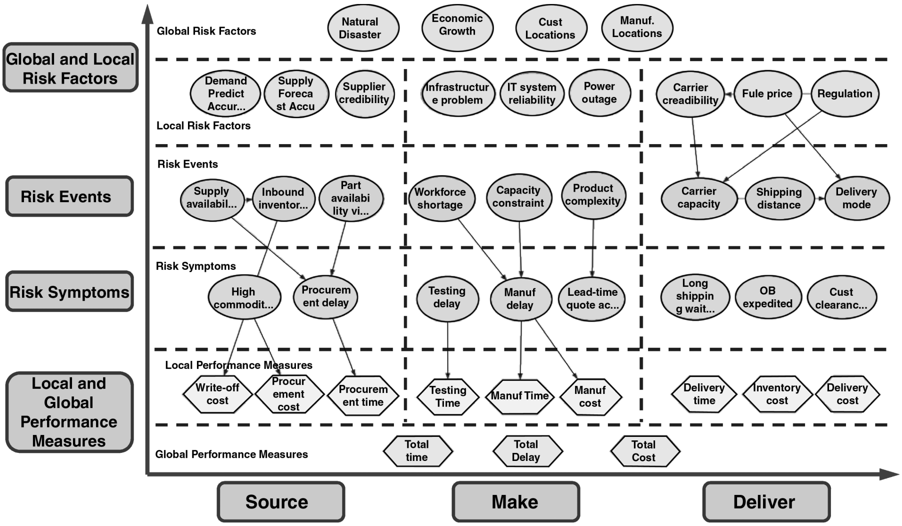


FIGURE 20.1 **Risk categorization framework.**

values. To address the problem, we define a set of performance measure variables that provide numeric valuations of the different business outcomes defined by the joint states of the risk events and symptoms.

20.3.2 STANDARD BUSINESS PROCESSES

The second dimension consists of the individual business processes. SCOR is a process reference model developed by the Supply Chain Council as the cross-industry standard for supply chain management. It is defined to describe the business activities of all phases of an enterprise. As a result, SCOR enables users to address, improve, and communicate supply chain management practices within and between all companies and industries.

The SCOR model defines a hierarchical structure among the individual business processes within a supply chain. For example, the first level contains five standard business processes: *plan*, *source*, *make*, *deliver*, *return*. Each of these processes contains more specific business sub-processes. We use the hierarchical structure of the business processes as the second dimension in the categorization, which facilitates the development of modularized design of the risk models.

20.3.3 A TWO-DIMENSIONAL RISK CATEGORIZATION FRAMEWORK

Figure 20.1 shows our proposed two-dimensional risk categorization. The vertical dimension defines the risk variable categories, including *risk factors*, *risk events*, *risk symptoms*, and *performance measures*. The horizontal dimension includes the various *business processes*. We can further introduce hierarchical structures within this categorization. In particular, we can decompose the business processes into more specific processes. *Global risk factors* are defined for the whole business process, and *local risk factors* are defined for each individual business processes. Similarly, we can also define *global performance measures* for the whole process and *local performance measures* for the individual processes.

The proposed categorization framework provides a structured approach for integrating the multiple information sources available to risk management, including historical data, business process standards, and expert knowledge. *Historical data* can be used to extract the risk variables, *business process standards* are used to define the hierarchical structures of the risk matrix, and *expert knowledge* can be utilized to categorize the risk variables by placing them in the appropriate slot in the matrix.

20.3.4 HIERARCHICAL CAUSAL STRUCTURE

The two-dimensional categorization framework provides an effective way to define the risk environment of a domain with a well-defined structure. First, the vertical categorization provides intuitive causal layers among the variables. Risk factors affect risk events, and risk events affect risk symptoms, but not the other way around. We also allow interaction within each of the categories, that is, some risk factors can affect other risk factors; similarly for risk events and risk symptoms. Also, the hierarchical business process structure provides a natural way to decompose the domain into relatively independent modeling problems. The global risk factors affect all the local risk factors and risk events but local risk factors can only affect the local risk events in the same process. The interaction between the different individual business processes are captured by the risk factors from higher level business processes. For example, when a risk event of a business process is observed, it changes the likelihood of the risk factors affecting this risk event, which in turn changes the likelihood of related risk events of other business processes. The effects of risk events and risk symptoms from different business processes are reflected through the local performance measures, and the local effects are combined together through the global performance measures. Figure 20.2 intuitively captures the *interaction flow.*

20.4 Risk Quantification through Bayesian Learning

The main goal of this chapter is to develop a framework for developing an end-to-end risk model for a supply chain by utilizing multiple information sources, including *business process standards*, *heterogeneous operational data*, and *expert knowledge* based on the risk categorization framework proposed in the last section. In particular, we will build risk models based on Bayesian networks (Pearl 1988). A Bayesian network provides a graphical representation of a domain in which nodes represent the random variables and directed arcs represent the uncertain

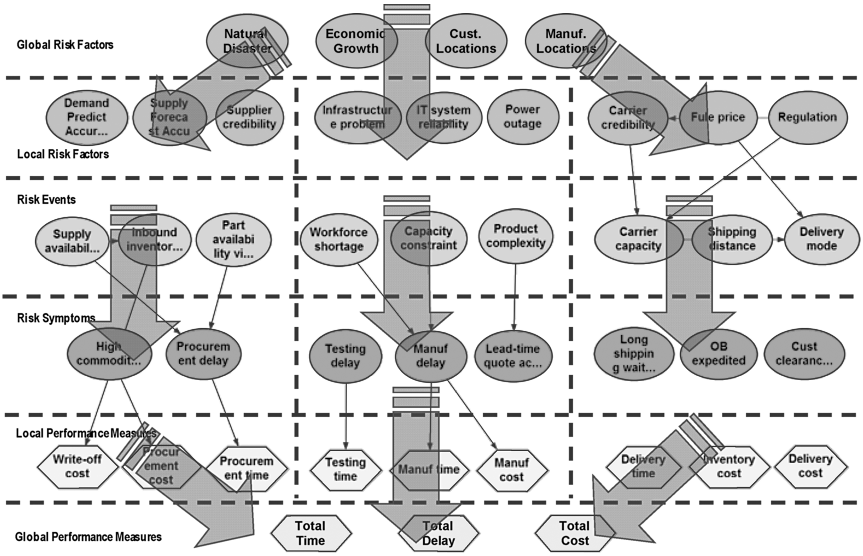


FIGURE 20.2 Interaction flow among the risk variables.

dependence relations among the variables. Although the directionality of an arc does not necessarily represent the causal direction between the two end variables, the arcs that do follow causal directions are more intuitive for humans and typically produce Bayesian networks with more succinct graphical structures. The categorization framework would enable us to learn a well-structured causal risk model.

20.4.1 LEARNING HETEROGENEOUS RISK MODEL

We propose to use the risk categorization framework presented in the last section to facilitate the learning of Bayesian risk models. Expert knowledge is only needed for classifying the data elements into different categories. Learning algorithms are then applied to learn the relations among the variables under the following two main constraints enforced by the categorization. First, in the order of global risk factors, local risk factors, risk events, risk symptoms, local performance measures, and global performance measures, arcs between these categories can only go from lower to higher categories while allowing arcs within the categories.

No arc is allowed to skip categories except that arcs are allowed between global risk factors to risk events. Second, arcs are allowed between the risk variables of the same category in each individual business process and no arcs are allowed between the risk variables from different business processes.

Utilizing the above constraints, we can apply any approach for learning Bayesian networks to learn the risk model. There are two main families of learning approaches: score-based methods and constraint-based methods.

20.4.2 SCORE-BASED LEARNING METHODS

Many approaches formulate learning Bayesian networks as an optimization problem. The idea is to define a quality measure, $Q(M|D)$ over different network structures and apply various search techniques to find one structure M to optimize the score, so called *local score metric-based methods*. For example, suppose we try to learn a model M from dataset D with N records. Let $r_i (1 \leq i \leq n)$ be the cardinality of variable x_i. Let q_i be the cardinality of the parent set of x_i in M, that is:

$$q_i = \prod\nolimits_{x_j \in pa(x_i)} r_j$$

When the parent set is empty, q_i is equal to *1*. We also let N_{ij} $(1 \leq i \leq n, 1 \leq j \leq q_i)$ be the number of records in D for which $pa(x_i)$ takes its jth value and N_{ijk} $(1 \leq i \leq n, 1 \leq j \leq q_i, 1 \leq k \leq r_i)$ be the number of records for which $pa(x_i)$ takes its jth value and x_i takes its kth value. Therefore:

$$N_{ij} = \sum_{k=1}^{r_i} N_{ijk}$$

Then the *entropy metric* $H(M,D)$ (Shannon 1948) of a network structure and dataset is defined as:

$$H(M, D) = -N \sum_{i=1}^{n} \sum_{j=1}^{q_i} \sum_{k=1}^{r_i} \frac{N_{ijk}}{N} \log \frac{N_{ijk}}{N_{ij}}$$

The number of parameters K as:

$$K = \sum_{i=1}^{n} (r_i - 1) q_i$$

Another metric called, AIC *metric* $Q_{AIC}(M,D)$ (Akaike 1974) of a Bayesian network structure M for a dataset D is defined as:

$$Q_{AIC}(M, D) = H(M, D) + K$$

MDL (minimum description length) (Rissanen 1978) metric $Q_{MDL}(M,D)$ is defined as:

$$Q_{MDL}(M, D) = H(M, D) + \frac{K}{2} \log N$$

The main difference between AIC and MDL is that they use different penalty terms to penalize the complexity of the learned structures.

There are yet some other approaches that measure the quality of a Bayesian network based on its prediction performance on a given data set, (e.g., classification accuracy). These are so called *global score metric-based methods*.

Given the above quality measures, we can apply search algorithms to look for a high-quality network structure. It has been shown that Bayesian network learning is extremely difficult as the search space is exponentially large, which is why early research focused on local search-based methods. Several popular greedy search algorithms have been developed. K2 algorithm (Cooper and Herskovits 1992) assumes that a total ordering of the variables is given and the directions of the arcs must follow the ordering. Then the algorithm greedily selects best arcs to add to an initially empty structure. We can also apply Markov chain Monte Carlo methods and simulated annealing to stochastically search for a high-scoring network (Heckerman et al. 1995).

20.4.3 CONSTRAINT-BASED LEARNING METHODS

Many other learning methods are based on *conditional independence tests*. These methods build on the same basic idea that the dependence and conditional independence relations between the variables can be determined using statistical testing methods. The directions of the arcs are assigned so that they satisfy the conditional independencies identified from the data. The PC algorithm (Spirtes et al. 1993) is such a method. It first creates an undirected graph based on the results of pair-wise independence testing. Then, it thins the model by sequentially removing edges with zero-order conditional independence relations, with first-order conditional independence relations, and so on. The *greedy thick thinning* algorithm (Cheng et al. 1997) first creates a draft model by computing pair-wise closeness measures based on mutual information. After that, the algorithm adds arcs when the pairs of nodes are not conditionally independent given conditioning variables. Finally, each arc is revaluated and will be removed if the two end nodes are independent.

In this chapter, we focus on adapting the greedy thick-thinning algorithm to create a model that obeys the constraints we have defined. The risk categorization framework uses expert knowledge to provide prior knowledge over the network structure, which will significantly reduce the size of the model space and makes it more likely to find a high-quality structure. The learning algorithm works largely in the same manner except that constraints are taken into account when adding or deleting arcs.

20.4.4 BAYESIAN INFERENCE FOR RISK ANALYSIS

Bayesian inference techniques allow reasoning the posterior probabilities of certain variables given observations on other variables. Let e be the observed states of a set of variables E, and X be the variable that we are interested in, and Y be all the other variables. We can calculate the posterior probability of X given that we observe e as follows:

$$P(X|E = e) = \sum_{Y} P(X, Y|E = e).$$

The junction tree algorithm (Lauritzen et al. 1988) allows posterior probabilities for all the unobserved variables be computed at once. Based on the reasoning direction, we can classify such inference tasks into predictive inference, where we reason from causes to effects, and diagnostic inference, where we reason from effects to causes.

With the risk mitigation strategies and performance measures defined, we can also test the sensitivity of the performance to risk-mitigation strategies or root causes. Suppose we want to test the sensitivity of performance measure M to risk mitigation strategy D, given observations e. We first turn off all the other risk mitigation strategies to single out D. We then systematically set D to its different states, which results in different probability distributions over unobserved variables X. We can use each distribution to compute the expected utility for action $D = d$ as follows:

$$EU_M(D = d|E = e) = \sum_X P(X|E = e, D = d)U_M(x)$$

Then, we can use the *difference* between the minimum and maximum of these expected utilities and/or their standard deviations to measure the sensitivity of M to D given e.

If the risk model is too complex, or if we want to get a distribution and not just a single number on the expected utility, we can apply Monte Carlo simulation techniques to estimate the utility distributions for different states of the mitigation strategy. If e is empty, *forward Monte Carlo simulation* can be performed. The idea is to sample each random variable given its parents according to a topological order of the model. Once all the random variables are sampled, a total utility can be computed by summing the individual values of all utility variables. We can repeat the above sampling procedure many times in order to obtain a set of expected utilities, which can be used to estimate a utility distribution. However, if e is not empty, *importance sampling* has to be used instead. The simplest strategy for performing importance sampling in Bayesian networks is a method called *likelihood weighting* (Fung and Chang 1989), in which we still sample the model using a topological order. However, whenever we encounter an observed variable, we simply use its observed state as the sample. This strategy obviously introduces bias because we are not sampling from the correct posterior distribution. We need to correct the bias by assigning *weights* to the samples. We can adapt likelihood weighting to solve an influence diagram. Here we derive the formula:

$$\begin{aligned}
&EU_M(D = d|E = e) \\
&= \textstyle\sum_X P(X|E = e, D = d)U_M(x) \\
&= \frac{1}{P(E = e|D = d)} \sum_X P(X, E = e|D = d)U_M(x) \\
&= \frac{1}{P(E = e|D = d)} \sum_X P(E = e|X, D = d)U_M(x)P(X|D = d).
\end{aligned}$$

Therefore, we can sample from $P(X|D{=}d)$ as we did before. The bias of each sample is corrected by multiplying the utility value with $P(E{=}e|X, D{=}d)$. We can repeat the process many times to get a set of weighted samples and use the *average of the weighted utilities* to estimate the expected utility as follows:

$$\begin{aligned} &EU_M(D = d|E = e) \\ &\approx \frac{1}{N} \cdot \frac{1}{P(E = e|D = d)} \sum_{x_i} P(E = e|x_i, D = d)U_M(x_i), \end{aligned}$$

Note that $x_i \propto P(X|D{=}d)$ and N is the number of samples. $P(E{=}e|D{=}d)$ can be estimated as follows:

$$\begin{aligned} &P(E = e|D = d) \\ &= \sum_X P(X, E = e|D = d) \\ &= \sum_X P(E = e|X, D = d)P(X|D = d) \\ &\approx \frac{1}{N} \sum_{x_i} P(E = e|x_i, D = d). \end{aligned}$$

Hence, we can use the same set of samples to estimate $P(E{=}e|D{=}d)$. The final estimator is asymptotically unbiased.

20.5 Case Study: Risk Modeling for a Global Supply Chain

In this section, we present a case study where we apply the proposed risk framework to analyze the delivery process of a server computer manufacturer's global supply chain.

This supply chain delivery process makes a good test case for evaluating the approach due to its scale and the availability of operational data.

The global supply chain we study offers about 50,000 unique configurations that include Intel-based servers, Blades, Blade centers, Racks, and Linux clusters. The company operates several manufacturing sites across the Americas, Asia, and Eastern Europe. The products are sold to customers all over the world. The delivery process includes three different routes to market:

1. Direct. Products are sold directly to end customers through e-commerce and telemarketing. This constitutes about 20% of the total volume.
2. Value-added resellers (VARs). Products are sold to VARs, who then sell the products to the end customers. This constitutes about 10% of the total volume. Distributors. Products are sold to large distributors, who sell the products to the VARs, who, in turn, sell the products to the end customers. This constitutes about 70% of the total volume.

When a customer places an order, a designated manufacturing site is selected to fulfill the order based on the customer location, products ordered, and the available manufacturing capacity at the different manufacturing sites. Customers usually request a desired delivery date when placing an order. Using the order-entry date and the current order pipeline, shipment and delivery dates are determined. The committed shipment and delivery dates depend on multiple factors:

- Whether the order contains standard products or highly customized product configurations.
- Whether the order requires bundling of multiple products.
- Whether the order must be expedited.

The requested products are assembled at the designated manufacturing facility. Based on the time of completion of manufacturing and potential for shipment consolidation, the order is subsequently shipped through a carrier who provides the best service given the delivery date requirements.

The metrics that are tracked include: the time between the order dropped on manufacturing and the order being shipped, tardiness with respect to the committed shipment date, the time between order shipment and order delivery, and tardiness with respect to the committed delivery date. The on-time performance metrics are tracked very closely and have high targets (e.g., 95% of shipments must be delivered on-time).

20.5.1 DATA PROCESSING AND LEARNED RISK MODEL

We collected several months' worth of historical supply chain delivery data on shipments that cover all products, all manufacturing facility locations, and all customer deliveries, including those to business partners, resellers, and end customers. Data elements include geographical details such as manufacturing plant locations, customer locations; products ordered along with order quantities,; order weight; order volume and transportation mode; order entry dates; requested, committed, and actual dates for shipment and delivery; and other details such as tied vs. untied orders, configured vs. standard products, and delay reasons.

In order to develop a Bayesian network model, we cleansed and transformed the operational supply chain data as follows. First, clearly irrelevant data elements or redundant attributes were deleted. Second, we mapped several qualitative variables to higher levels of granularity such as mapping products to product families and mapping customers to customer types. Third, we used percentile-based methods to estimate new data. For example, in order to estimate demand levels, we used the shipping quantities by product and day to estimate the monthly demand distribution and then used the 5th and 95th percentiles to discretize the demand levels. Finally, in order to be able to perform a multiple-fault risk analysis, we translated each cause/category into independent risk factors.

Next, we identified the risk variables and used the risk-categorization framework to classify them into three categories: risk factors, risk events, and risk symptoms. In addition, we defined performance measures to assign quantifiable business value to the different outcomes of risk events and symptoms.

After preprocessing the data, we applied the Bayesian learning method described earlier to learn the risk model (Figure 20.3). After we learned the probabilistic component (for risk variables), we also added the utility nodes that defined the values for different business scenarios. Several risk mitigation strategies were also defined to mitigate the impact of some of the root causes on the performance measures.

20.5.2 RISK MODELING AND ANALYSIS

Based on the Bayesian framework proposed earlier, we developed a model to perform risk analysis based on Bayesian inference techniques. The model essentially represents a joint probability distribution of all the risk variables, the business value of different outcomes, and the effect of risk mitigation strategies on the model. The model allows us to perform many kinds of risk analyses, including *risk prioritization*, *diagnosis*, *mitigation strategy evaluations,* and others. The capabilities of the model are summarized in Table 20.2.

In the following subsections, we describe each of the above risk analysis capabilities in details.

TABLE 20.2 Capabilities of the Risk Modeling and Analysis Techniques

Input	Capability	Output
• Historical data to assess relations and impact • Expert knowledge • Online supply chain operational data	• Model risk factors, risk events, risk symptoms, performance measure network • Model risk dependencies and conduct multivariate Bayesian analysis • Process supply chain data for risk diagnostics • Generate values for prioritization of risk factors • Perform scenario analysis and impact analysis • Perform Monte Carlo simulation for risk factors • Model and evaluate risk mitigation strategies	• Risks—likelihood and impacts • Risk mitigation options • Business impacts

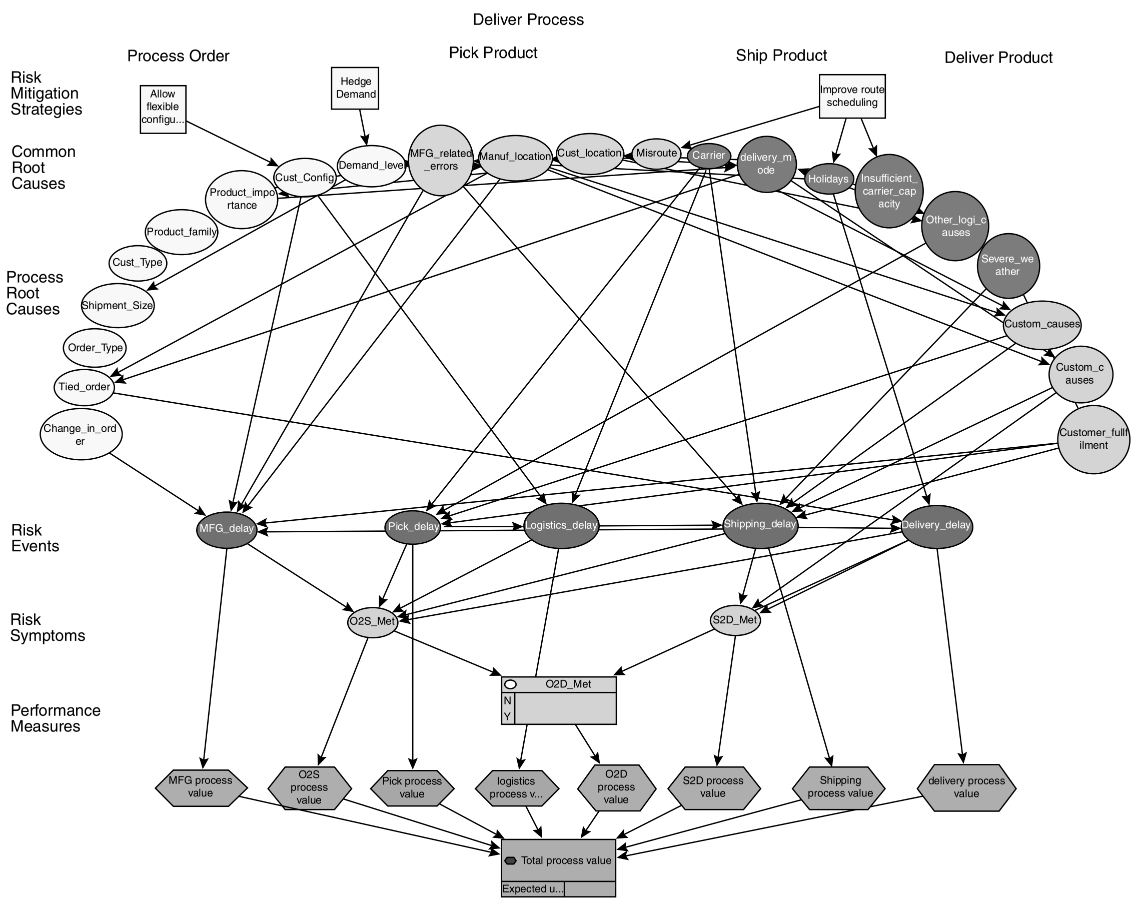


FIGURE 20.3 Bayesian risk model for global supply chain delivery process.

20.5.2.1 Analyzing the Likelihood of Risk Factors. The likelihood of risk factors, risk events, and risk symptoms is computed based on the posterior probability distributions of the variables. By analyzing the causal relationships among various risk factors, we can establish and quantify the dependency between these factors, and, therefore, identify the root causes of risk events.

There are two types of risk factors: those for which normal and abnormal states are well defined, and those for which that is not the case. For instance, weather can be a risk factor with its normal state and abnormal state defined, based on whether a severe weather condition occurs for the given time period. However, other risk factors such as the manufacturing location have no predetermined abnormal states.

The output from this analysis includes a list of risk variables ranked based on the likelihood of their occurrence. Typically, the ranking is provided for the variables with well-defined abnormal states and is based on the largest probability of an abnormal state. The distribution of the possible states for any given risk factor is also computed based on historical data. Using the results of this analysis, we can quickly identify the risk factors that have the highest likelihood of occurrence as illustrated in Figure 20.4. Furthermore, the causal relationships represented in the Bayesian network allow us to track root causes easily.

20.5.2.2 Analyzing the Impact of Risk Factors. In this subsection, we show how we can analyze the impact of individual risk factors, events, or symptoms on a given performance measure. As illustrated by Figure 20.5, this is achieved by setting the selected risk variable to different states and then calculating the expected values of the performance measure for each state. We can also use Monte Carlo simulation to obtain the distribution of the performance measure under randomly sampled settings of the risk variable. The histogram of the results can also be generated to approximate the distribution of the performance measure. The distribution allows us to perform risk assessment using a value-at-risk based calculation, which is better than the usual expected value calculation for risk factors with low frequency and high impact.

Combining the results of the likelihood and impact analyses, we can obtain a risk matrix as shown in Figure 20.6. This matrix provides a clear picture of the impact and likelihood of different risk factors. This can be used to highlight risk factors with high likelihood and high impact.

20.5.2.3 Analyzing the Effect of Mitigation Strategies. Next, we analyze the effect of a mitigation strategy on a given performance measure. To perform this analysis, we need to define the mitigation strategy as a decision node in the Bayesian network model. A mitigation strategy is expected to change the frequency and severity of certain risk factors. When historical data is available, it allows us to observe the changes to the Bayesian network parameters when the mitigation strategy is used. However, in most practical situations, such changes will have to be made based on expert knowledge.

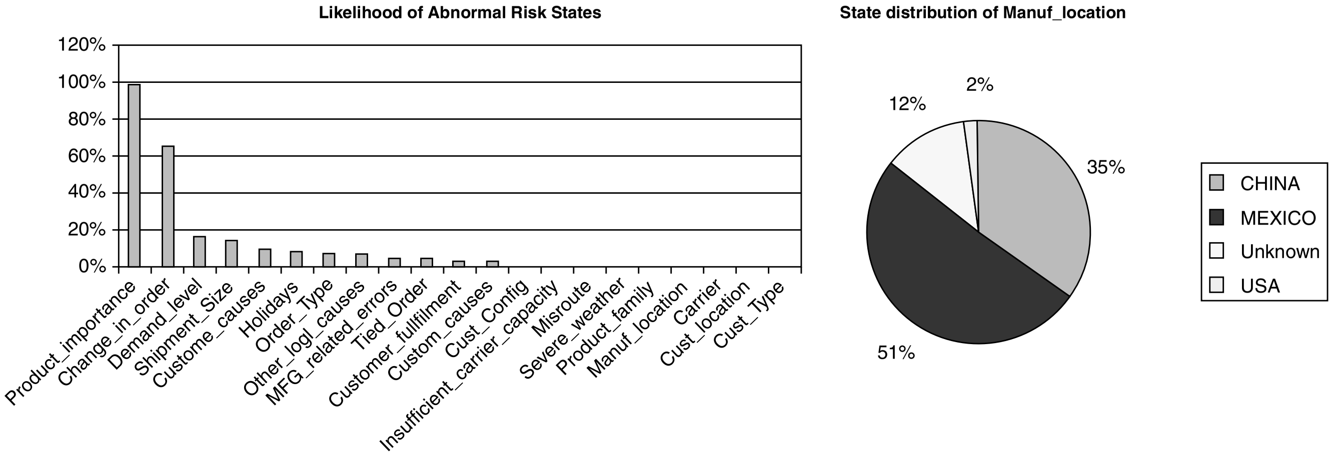


FIGURE 20.4 Analyzing the likelihood of risk factors: (a) likelihood by risk factor, (b) likelihood by state for selected risk factor.

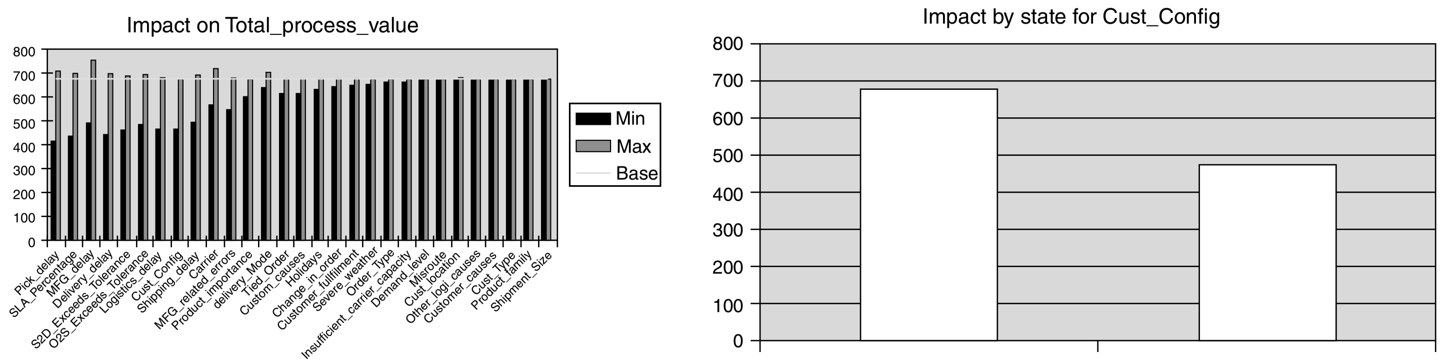


FIGURE 20.5 Analyzing the impact of risk factors: (a) impact by risk factor, (b) impact by state for selected risk factor.

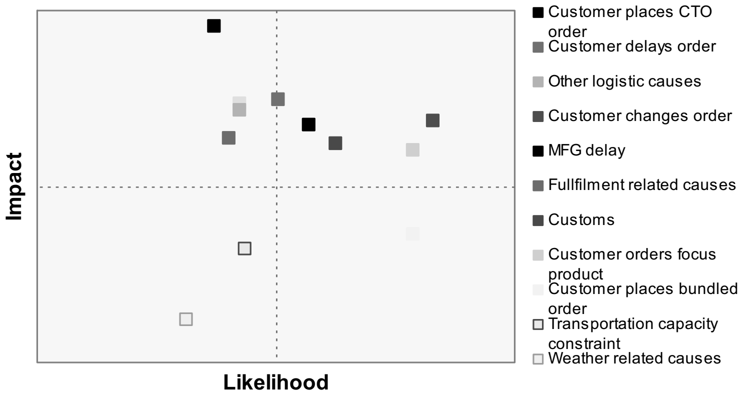


FIGURE 20.6 **Risk matrix with likelihood vs. impact.**

Once the new decision node is added to the Bayesian network, the performance measure can be computed as discussed in the above subsection for each state of the decision node. Furthermore, Monte Carlo simulations can be used to compute the distribution of the performance measure for "on" and "off" settings of the decision node. The difference between the two settings allows us to assess the impact of the mitigation strategy.

Using the same analysis for multiple mitigation strategies along with the cost of implementing these strategies, we can figure out the most cost-effective risk mitigation strategy for a given performance measure (see Figure 20.7 for illustration).

In some cases, we can use mathematical programming to identify the optimal decision. For such optimization problems, the objective is a utility function of the relevant performance measures and the constraints include the budgets for implementing the mitigation strategies.

20.5.3 USE CASES OF THE CASE STUDY

With the risk assessment and quantification techniques outlined in the previous section, we present several use cases developed from our case study in the context of the global supply chain.

20.5.3.1 Use Case 1: Identifying Key Root Causes by Examining Parent Nodes of Risk Symptoms. The objective of this use case is to identify the key root causes for the risk symptoms. There are multiple ways to accomplish this. In this chapter, we start from the risk symptoms and then trace back to their parent nodes using the utilities provided by the Bayesian network model.

Since the risk symptoms are directly associated with the performance measures, we can observe the risk symptoms. For example, one risk symptom in the global supply chain model is the time from order to ship exceeding the service-level target. Using the Bayesian network model shown in Figure 20.8, we can easily

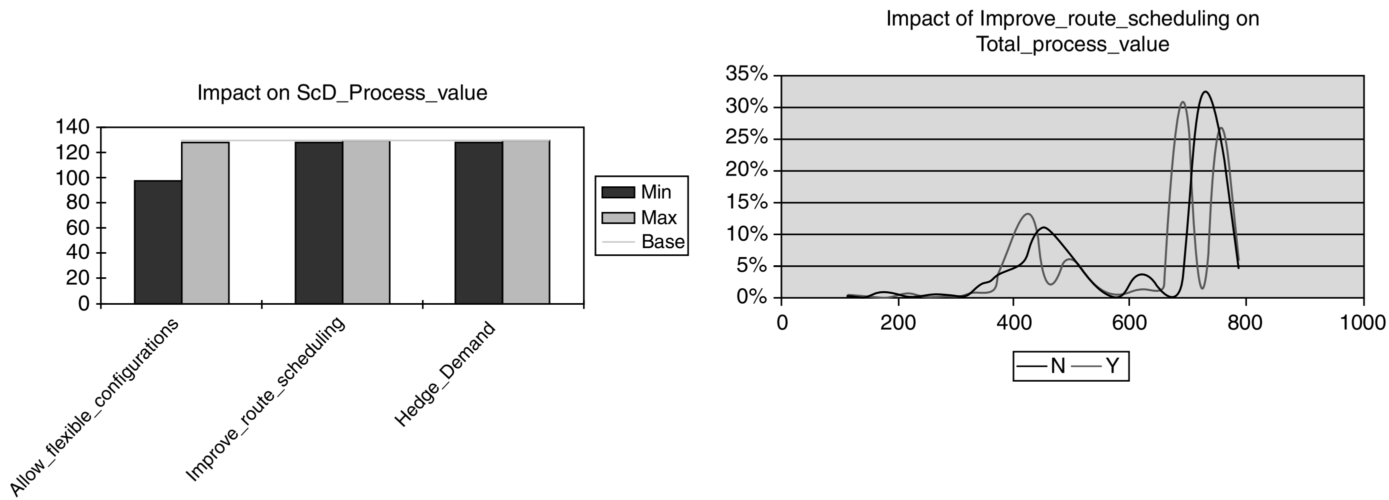


FIGURE 20.7 **Analyzing the impact of mitigation strategies: (a) impact analysis, (b) simulation analysis.**

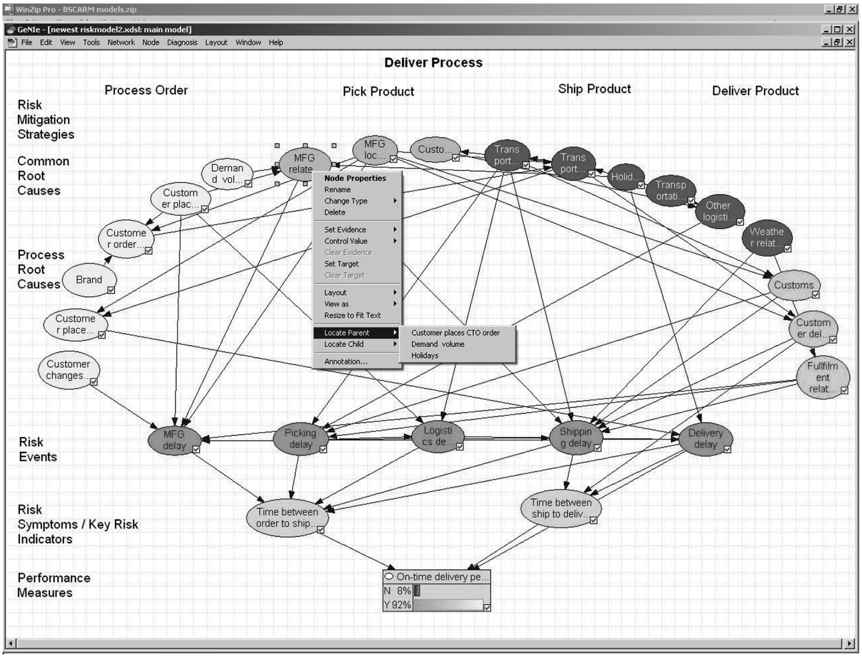


FIGURE 20.8 **Use Case 1: Identifying root causes.**

identify three risk factors (“customer places CTO order,” “demand volume,” and “holidays”) that are the causes for this risk symptom. We can then examine the likelihood and the business impact of these risk factors. The likelihood of occurrence for a risk factor is provided by the Bayesian network model as a probability distribution of the possible states for the risk factor. The impact of a risk factor can be determined by comparing the probability distributions of the risk symptom under different states of the risk factors.

20.5.3.2 Use Case 2: Identifying Key Root Causes by Examining the Root Nodes with Different Evidence Settings for Risk Symptoms. The approach described in Use Case 1 requires performing multiple influence analyses iteratively to trace the parent nodes. To pinpoint the root causes of a risk symptom, we can also fix the risk symptom level and observe the changes of risk likelihood at the root-cause level.

For example, in Use Case 1 we set “time from order-to-ship exceeds tolerance” to “yes” and observed the risk factors and their risk profiles under this new setting. By comparing them with the “as-is” risk profiles where without fixing any variables, we are able to see the difference in the risk profiles between the two settings. If one risk factor shows a bigger difference in terms of likelihood and impact under the two settings, we can reasonably assert that this particular risk factor has a higher impact on the risk symptom selected. On the other hand, if

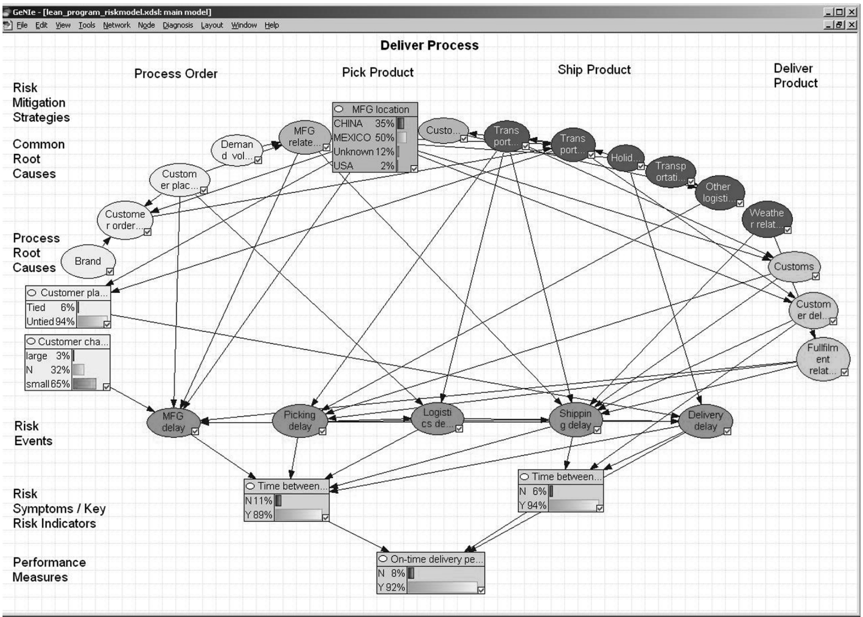


FIGURE 20.9 **Use Case 2: Identifying root causes by setting evidence at risk symptom.**

the difference is small or negligible, we can also conclude that the risk factor has a lower impact on the risk symptom.

In the example shown in Figure 20.9, the risk profile of "manufacturing location" shows a significant change before and after the evidence is set for the risk symptom "time from order-to-ship exceeds tolerance." When evidence is turned off (i.e., "as-is"), less than 45% of the total volume shipped to the U.S. is from Mexico. However, once the evidence is set, 55% of the total volume (of which their order-to-ship exceeds tolerance) shipped to the U.S. is from Mexico. This indicates that the manufacturing location has a significant impact on the order-to-ship cycle time.

20.5.3.3 Use Case 3: Assessing the Impact of risk Mitigation Strategies. In this use case, we provide an impact assessment for a risk mitigation strategy. We first add a decision node that represents the mitigation strategy in the Bayesian network model. This process usually involves additional learning using either historical data or expert knowledge to calibrate the model parameters. Once the decision node is added and model parameters are set, performing the impact analysis for a mitigation strategy is straightforward. A mitigation strategy is proposed for implementing a lean program at the Mexico plant that will reduce the manufacturing cycle time by two days. We observe from our numerical analysis that the overall on-time performance in the "to-be" scenario is improved by 11% over the "as-is" scenario.

20.6 Summary

Globalization and growing supply chain interdependence have introduced a heightened level of volatility and vulnerability that is unlikely to subside (Friedman, 2005). Uncertainty has become the norm. This new environment demands a smarter and more resilient supply chain (IBM 2008, 2009). As companies transition to the smarter supply chain of the future, the ability to bring together all key stakeholders of the extended supply chain as well as relevant regulations and policies and to facilitate joint planning, risk mitigation, and responses across all business functions and strategic decision making will be essential to a company's success (Lin et. al. 2002, Buckley et.al., 2005, Lin and Wang 2010).

In this chapter, we presented a new framework for integrated supply chain risk modeling that utilizes multiple information sources including heterogeneous operational data, expert knowledge, and hierarchical business processes based on a two-dimensional risk categorization of risk factors and business process standards. We applied constrained Bayesian learning techniques to enable the learning of a risk model that obeys constraints defined by the risk categories. We also described a case study based on the delivery process of a global supply chain. Our approach goes beyond prior work by addressing complex and heterogeneous data and hierarchical business processes in today's global supply chains. The resulting risk model allows analysts to identify areas of high risks in supply chain business processes, to diagnose risk factors contributing to abnormal observations, and to analyze the sensitivity of a supply chain's performance measures in relation to key risk factors. This new model also provides guidance for identifying risk-mitigation strategies to reduce an organization's risk exposure, as well as quickly responding to disruptive events immediately after they occur.

Future work will focus on connecting the Bayesian models with various financial measurements, and integrate them with supply chain operation and planning models. We are also exploring advanced learning algorithms that utilize the categorization framework to update risk models continuously with daily operational data. The resulting models will enable companies to adapt smarter capabilities to control supply chain risks, and manage their global extended supply chain more effectively.

REFERENCES

Aberdeen Group. 2009. Supply chain risks in the uncertain economy: grappling with changes. Aberdeen Group. Available at: http://www.aberdeen.com/Aberdeen-Library/6217/RB-supply-chain-risk.aspx. Accessed August 3, 2009.

Akaike H.1974. A new look at the statistical model identification. *IEEE Transactions on Automatic Control*,19 (6):716–723.

Associated Press. 2007. Sony profits hit hard by PlayStation 3 launch. January 30. Available at: http://www.foxnews.com/story/0,2933,248743,00.html. Accessed October 10, 2010.

Basu G et al. 2008. Supply chain risk management: A delicate balancing act, a multi-faceted view on managing risk in a globally integrated enterprise. White paper. IBM Global Business Services. Available at ftp://ftp.software.ibm.com/common/ssi/sa/wh/n/gbw03015usen/GBW03015USEN.PDF.

Bouckaert RR. 1993. Belief network construction using the minimum description length principle. Proceedings ECSQARU, Granada, Spain, November 8–10, 1993. pp 41–48.

Buckley, S, Ettl M, Lin G, Wang K. 2005. Intelligent business performance management—sense and respond value net optimization. In: Fromm H. An C., editors. *Supply Chain Management on Demand*. Springer, New York. pp. 309–337.

Cheng F, Gamarnik D, Jengte N, Min W, Ramachandran B. 2007. Modeling operational risk in business processes. Journal of Operational Risk, 2(2):73–99.

Cheng J, Bell D, Liu W. 1997. An algorithm for Bayesian belief network construction from data. Proceedings of AI & STAT'97. pp 83–90.

Chozick A. 2007. Toyota sticks by 'just in time' strategy after quake. *The Wall Street Journal*, July 24, p A2. Available at http://online.wsj.com/article/SB118518000701874802.html.

Cooper GF, Herskovits E. 1992. A Bayesian method for the induction of probabilistic networks from data. *Machine Learning,* 9:309–347.

Dao H, Torpy J. 2008. Supply chain risk management. Atlanta Supply Chain Plenary.

Deleris L, Cope E. 2007. Sources of variability in IBM's system X supply chain. IBM Research report. Yorktown Heights, New York: IBM T.J. Watson Research Center,

Deleris LA, Erhun F. 2011. Risk management in a supply network: A case study based on engineering risk analysis concepts. In: Kempf K, Keskinocak P, Uzsoy R, editors. *Handbook of Production Planning. Kluwer International Series in Operations Research and Management Science,* Kluwer Academic Publishers, New York.

Druzdzel MJ, van der Gaag LC. 2000. Building probabilistic networks: Where do the numbers come from? (guest editors' introduction). *IEEE Transactions on Knowledgement and Data Engineering*, 12(4):481–486.

Ezawa KJ, Singh M, Norton SW. 1996. Learning goal oriented Bayesian networks for telecommunications risk management. Proceedings of the 13th International Conference on Machine Learning (ICML), Morgan Kaufmann. pp 139–147.

Feldman S, Gray W, King A, Perret R, Shi D. 2007. Apparatus, system and method for measuring and monitoring supply chain risk. US patent 7,246,080 B2. July.

Friedman T. 2005. *The World is Flat*: Farrar, Straus & Giroux, New York.

Fung R, Chang KC. 1989. Weighing and integrating evidence for stochastic simulation in Bayesian networks. In: M. Henrion, R. Shachter, L. Kanal and J. Lemmer, editors. *Uncertainty in Artificial Intelligence 5*. New York: Elsevier Science Publishing. pp 209–219.

Heckerman DE, Geiger D, Chickering DM. 1994. Learning Bayesian networks: The combination of knowledge and statistical data. Proceedings of the Tenth Conference on Uncertainty in Artificial Intelligence, Morgan Kaufmann, July 29–31, 1994, Seattle, WA. pp 293–301.

Hirotugu A. 1974. A new look at the statistical model identification. *IEEE Transactions on Automatic Control,* 19(6):716–723.

IBM. IBM 2008 CEO study report—enterprise of the future. 2008. Available at: http://www.ibm.com/ibm/ideasfromibm/us/ceo/20080505/, May 25.

IBM. IBM Smarter Planet. 2009. Available at: http://www.ibm.com/smarterplanet, November 1.

IBM. IBM Global Chief Supply Chain Officer Study. 2009. IBM. Available at: http://www-03.ibm.com/press/us/en/pressrelease/26774.wss.

Kao HY, Huang CH, Li HL. 2005. Supply chain diagnostics with dynamic Bayesian networks. *Computers and Industrial Engineering,* 49:339–347.

Kraaijeveld P, Druzdzel MJ. 2005. GeNIeRate: An interactive generator of diagnostic Bayesian network models. In: Working Notes of the 16th International Workshop on Principles of Diagnosis (DX-05). June 1–3, Monterey, California. pp 175–180.

Lam W, Bacchus F. 1994. Learning Bayesian belief networks, an approach based on the MDL principle. C*omputational Intelligence*, 10(4):269–293.

Lauritzen SL, Spiegelhalter DJ. 1988. Local computations with probabilities on graphical structures and their application to expert systems. *Journal of the Royal Statistical Society*, Series B50:157–224.

Lin G, et al. The sense and respond enterprise. 2002. *OR/MS Today*, 29(2):34–39.

Lin G, Wang K. Sustainable globally integrated enterprise. 2010. In: Sodhi MS, Tang CS, editors. *A Long View of Research and Practice in Operations Research and Management Science: The Past and the Future.* New York: Springer. pp 195–217.

Miller C, Frankel Y, Rosenkrantz N. 2005. System and method for risk detection reporting and infrastructure. US patent 2005/0021360 A1.

Naim MM, Childerhouse P, Disney SM, Towill DR. A supply chain diagnostic methodology: Determining the vector of change. 2002. *Computers and Industrial Engineering*, 43:135–157.

Norrman A., Jansson U. 2004. Ericsson's proactive supply chain risk management approach after a serious sub-supplier accident. *International Journal of Physical Distribution & Logistics Management*, 34(5):434–456.

Pearl J. 1988. *Probabilistic Reasoning in Intelligent Systems: Networks of Plausible Inference.* San Francisco: Morgan Kaufmann Publishers.

Pender K. 2001. Write-offs remove excess inventory from books. *San Francisco Chronicle.* May 8. Available at- http://articles.sfgate.com/2001-05-08/business/17596938_1_offs-restructuring-charges-market-value. Accessed October 10 2010.

Ray BK, Apte C, McAuliffe K, Cope E. 2007. Harnessing uncertainty: The future of risk analytics. Research report. Yorktown Heights, New York: IBM T.J. Watson Research Center.

Rigby B. Boeing profit up on plane sales. 2008. Reuters News. Available at http://www.reuters.com/article/idUSN3044925220080130. Accessed October 10, 2010.

Rissanen J. 1978. Modeling by shortest data description. *Automatica,* 14:465–471.

Shannon CE. 1948. A mathematical theory of communication. *Bell System Technical Journal,* 27:379–423, 623–656.

Sodhi M, Tang CS. 2009. Economic recovery through supply chain reasoning. *OR/MS Today.* June. Available at http://www.lionhrtpub.com/orms/orms-6-09/frviewpoint.html. Accessed October 24 2010.

Spirtes P, Glymour G, Scheines R. 1993. *Causation, Prediction and Search.* Berlin: Springer.

Supply Chain Council. 2009. Supply-Chain Operations Reference Model (SCOR). Available at: http://www.torconsulting.com/downloads/scor7_overview.pdf.

Tang CS. 2006. Perspectives in supply chain risk management. *International Journal of Production Economics*,103(2):451–488.

Treur CWC. 2008. A model for supply chain risk analysis and management. Master's thesis. University of Groningen, Groningen, the Netherlands.

Watthayu W, Peng Y. 2004. A Bayesian network based framework for multi-criteria decision making. In: multi-criteria decision making (MCDM) conference, 2004. Whistler, British Columbia, Canada. August 6–11, 2004. Available at: http://ebiquity.umbc.edu/get/a/publication/235.pdf.

Wu T, Blackhurst J, Chidambaram V. 2006. A model for inbound supply risk analysis. *Computers in Industry,* 57:350–365.

Index

Handbook of Integrated Risk Management in Global Supply Chains, First Edition. Edited by Panos Kouvelis, Lingxiu Dong, Onur Boyabatli, and Rong Li.