Supply Chain Management: Models, Applications, and Research Directions

Applied Optimization

Volume 62

Series Editors:

Panos M. Pardalos
University of Florida, U.S.A.

Donald Hearn
University of Florida, U.S.A.

The titles published in this series are listed at the end of this volume.

Supply Chain Management: Models, Applications, and Research Directions

Edited by

Joseph Geunes

Department of Industrial and Systems Engineering,
University of Florida, Gainesville

Panos M. Pardalos

Department of Industrial and Systems Engineering,
University of Florida, Gainesville

and

H. Edwin Romeijn

Department of Industrial and Systems Engineering,
University of Florida, Gainesville

KLUWER ACADEMIC PUBLISHERS

NEW YORK, BOSTON, DORDRECHT, LONDON, MOSCOW

eBook ISBN: 0-306-48172-3
Print ISBN: 1-4020-0487-7

Created in the United States of America

Visit Springer's eBookstore at: http://ebooks.springerlink.com
and the Springer Global Website Online at: http://www.springeronline.com

Contents

Preface

This work brings together some of the most up to date research in the application of operations research and mathematical modeling techniques to problems arising in supply chain management and e-Commerce. While research in the broad area of supply chain management encompasses a wide range of topics and methodologies, we believe this book provides a good snapshot of current quantitative modeling approaches, issues, and trends within the field. Each chapter is a self-contained study of a timely and relevant research problem in supply chain management. The individual works place a heavy emphasis on the application of modeling techniques to real world management problems. In many instances, the actual results from applying these techniques in practice are highlighted. In addition, each chapter provides important managerial insights that apply to general supply chain management practice.

The book is divided into three parts. The first part contains chapters that address the new and rapidly growing role of the internet and e-Commerce in supply chain management. Topics include e-Business applications and potentials; customer service issues in the presence of multiple sales channels, varying from purely Internet-based to traditional physical outlets; and risk management issues in e-Business in B2B markets.

The second part of the book deals with problems of coordinating the activities of different players within the supply chain. Topics include the impact and management of uncertainty when selling perishable seasonable products through mechanisms such as: advance booking discounts in the case of long replenishment lead-times; partial quick response policies when a second ordering opportunity is available; and a stochastic programming based decision support system. Other topics included are the effect of revenue sharing on the purchasing behavior of a vendor; supply chain contracting and coordination with shelf-space-dependent demand; a fee-setting model to decide a manufacturer's compensation scheme for the services provided by its independent distributors; tactical distribution planning with resource acquisition and deployment decisions; and mechanisms for controlling retail store-order variability to improve supply chain performance.

Finally, the third part focuses on models and applications for supply chain planning and design. Topics explored include: the design of global facility networks; a planning model for multiple products manufactured across multiple manufacturing facilities sharing similar production capabilities; models for evaluating logistics costs in a global supply chain in the aviation industry; supply chain models in the forest industry; and a study on the benefits of information sharing in the supply chain.

This book can serve as a valuable reference for researchers in supply chain management as well as a reference text book for a graduate level reading course.

All chapters in this book were thoroughly refereed by two anonymous referees. We would like to take this opportunity to thank the authors of the chapters, the referees, as well as several Ph.D. students at the Department of Industrial and Systems Engineering at the University of Florida, for their efforts. We would like to give special thanks to: Zuo-Jun "Max" Shen, Burak Eksioglu, Sandra Duni Eksioglu, Olga Perdikaki, Kevin Taaffe, M. Bayram Yildirim, and Joongkyu Choi.

JOSEPH GEUNES, PANOS PARDALOS, AND EDWIN ROMEIJN

I

THE ROLE OF THE INTERNET AND E-COMMERCE IN THE SUPPLY CHAIN

Chapter 1

SUPPLY CHAIN INTEGRATION OVER THE INTERNET

Hau L. Lee
Graduate School of Business
Stanford University
haulee@stanford.edu

Seungjin Whang
Graduate School of Business
Stanford University
whang-jin@gsb.stanford.edu

Abstract The Internet has emerged as the most cost effective means of driving supply chain integration. We define e-Business as the marriage between the Internet and supply chain integration. We divide various forms of e-Business applications into three categories–e-Commerce, e-Procurement, and e-Collaboration. e-Commerce helps a network of supply chain partners to identify and respond quickly to changing customer demand captured over the Internet. e-Procurement allows companies to use the Internet for procuring direct or indirect materials, as well as handling value-added services like transportation, warehousing, customs clearing, payment, quality validation, and documentation. e-Collaboration facilitates coordination of various decisions and activities beyond transactions among the supply chain partners over the Internet. This article studies various e-Business applications and discusses the potential of e-Business for building intelligence and optimization.

1. Introduction

Supply chain management (SCM) shifts the unit of analysis from a plant, a warehouse or a company to the entire supply chain. Since a supply chain typically spans over multiple companies, SCM particularly highlights the importance of cross-enterprise coordination – in the name

J. Geunes et al. (eds.),
Supply Chain Management: Models, Applications, and Research Directions, 3–17.

of supply chain integration. But supply chain integration requires a cost-effective information system that links multiple companies. This need can now be met by the Internet. We term this marriage between supply chain integration and the Internet as "e-Business" Thus, e-Business is here defined as executing front-end and back-end operations in a supply chain using the Internet. In a sense, e-Business is on the natural growth path of enterprise information systems that started with material procurement (MRP, material requirements planning), expanding to manufacturing (MRPII, manufacturing resource planning) and to intra-enterprise integration (ERP, enterprise resource planning). This article studies various e-Business applications and discusses the potential of e-Business for building intelligence and optimization.

2. Cross-docking Information Flows

The Internet is an electronic link that ties different entities. But the Internet is neither the first nor the only electronic link. For example, EDI (Electronic Data Interchange) on VAN (Value Added Network) is another electronic link that preceded the Internet age. But the Internet has many more advantages–it is based on open standards and grants universal access to a wide audience (anytime, anyplace, anyone, almost) at a lower cost. But most of all, the key power of the Internet is a new system architecture consisting of a hub and spokes. To underscore this point, consider a hypothetical company called Acme that manufactures a product and sells through a traditional three-tier distribution channel–resellers, wholesalers and distributors. Consider the information flows through the supply chain. A customer places an order with the reseller, who finds that it does not have enough stock in its own warehouse. The reseller places an order with its wholesaler. The wholesaler in turn may order from the distributor, and so forth. This way, order information flows sequentially through resellers, wholesalers, distributors, Acme, parts suppliers and logistics providers. At each step data are collected, batched, manipulated and transmitted multiple times, incurring cost and time delays.

Interestingly enough, this inefficiency in information flows is analogous to that often observed in material flows. In the traditional three-tier distribution system, finished goods flow from the plant to the manufacturer's warehouse, then to the distributor's warehouse, the wholesaler's warehouse, the retailer's warehouse, and finally to its stores. In the process, logistics cost accumulates due to additional handling, multiple handoffs, document processing, transportation, warehousing, and excessive inventories throughout the supply chain. In the case of the grocery industry, for example, this inefficiency previously resulted in over 120 days' supply of inventory and $30 billion of unnecessary cost (Kurt Salmon Associates Inc (1993)). One logistics solution is 'cross docking,' by which products delivered from multiple plants to multiple distribution centers (DCs), or from multiple DCs to multiple retail stores, are re-assorted and transported to the destinations without staying at a warehouse. In a typical cross-docking scenario, a group of inbound single-SKU (stock keeping unit) trucks arrive from various manufacturers to one side of the cross-dock facility. The content is unloaded onto the dock, sorted, and redistributed in smaller lots to outbound trucks waiting on the other side of the dock. When the operation is completed in a matter of hours, outbound multi-SKU trucks leave for stores. This way the 'time' dimension, along with its associated cost, is taken out of logistics.

One can extend and apply cross-docking to information flows to derive the 'information hub' where information is instantaneously processed and forwarded to all partners upon arrival. The information hub (see Bock (1998) and Lee and Whang (2000)) is a node in the data network where multiple organizations interact in pursuit of supply chain integration. It has the capabilities of data storage, information processing, and push/pull publishing. The overall network forms a hub-and-spoke system with the participants' internal information systems (i.e., ERP or other enterprise systems) being the spokes (see Figure 1.1). The information hub would be a web site or a central server on the Internet running an ERP system for the supply chain. Acme, if equipped with its front-end web interface and back-end supply chain integration operating through the information hub, represents a model of e-Business.

Continuing with our Acme example, suppose a reseller places a replenishment order to Acme on the web. This order is captured at the hub. Following certain agreed-upon protocols, this order information is pulled by, or pushed to, relevant supply chain partners – Acme, suppliers, logistics service providers, and financial institutions. The difference here is the real-time parallel processing of information, instead of sequential processing in the batch mode. Other interactions such as order tracking and forecast sharing can be accommodated by this architecture. The hub architecture offers a new paradigm of coordinating the activities of the supply chain from end customers to different supply chain partners. It makes the Internet distinct from other electronic links such as EDI over VAN.

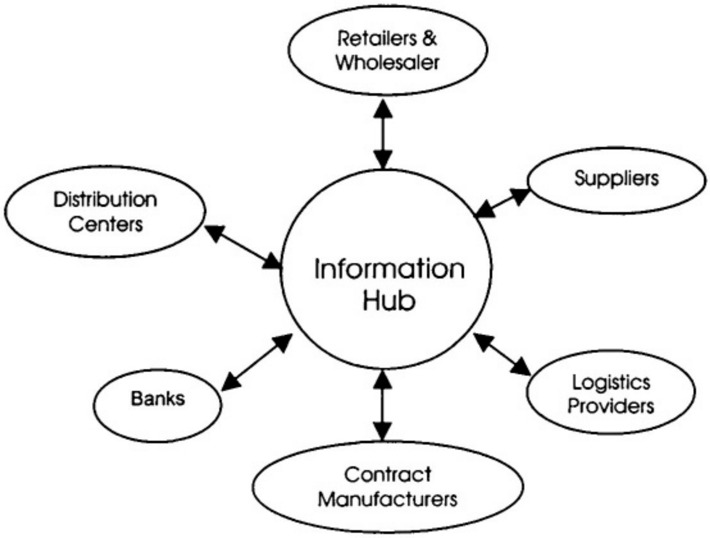


Figure 1.1. The Information Hub

3. Three Classes of e-Business Applications

For convenience of discussion, we divide e-Business applications to three classes – e-Commerce, e-Procurement and e-Collaboration. We first provide a brief description of each class, followed by a full discussion of each.

A. e-Commerce

When a customer places an order on the web, the order triggers a series of transactions throughout the supply chain. A speedy and accurate execution of a transaction is perhaps the most fundamental form of interaction among supply chain partners. In addition, a cross-enterprise system is required to track the status of an order with one call, no matter which partner is currently holding the order. Lastly, after-sales service should be also captured as a component of supply chain integration, since it requires continuing interactions with supply chain partners and end customers. The information hub can offer a natural platform to capture the order, coordinate the activities, track the order status and deliver after-sales service. Besides tracking and executing order flow activities, the information hub can also offer performance measures linked to the supply chain, such as lead time, quality and inventory turns. These applications are called e-Commerce.

B. e-Procurement

Modern manufacturing requires flexibility due to stiff competition, fast changing customer preferences, shortening product life cycle and product variety proliferation. Along with dynamic capacity allocation, efficient material procurement forms a pillar to support flexible manufacturing. The ERP system addresses the needs of material planning, but execution of procurement is outside the scope of a typical ERP system. The actual procurement process reaches beyond the enterprise level and requires extensive interactions with suppliers. The hub architecture again offers a favorable setting in which numerous buyers and sellers can find each other and transact according to some pre-specified protocols (governed by the marketplace or traders' internal rules). This we call e-Procurement.

C. e-Collaboration

Supply chain integration implies more than the traditional arm's length relationship based on market transactions among its partners. It often involves sharing of information and knowledge that used to be thought as proprietary or even strategic. Examples of the information being shared are sales data, inventory status, production schedule, promotion plans, demand forecasts, shipment schedule, and new product introduction plans. Even further, supply chain partners can make joint decisions based on combined information and knowledge (Lee (1998)). Without information sharing or joint decisions, for example, orders to the supplier tend to have larger variance than sales to the buyer, and the distortion propagates upstream in an amplified form – the phenomenon called the "bullwhip effect" (Lee, Padmanabhan and Whang, (1997)). One way of mitigating the bullwhip effect is for partners to delegate the inventory decision either to the vendor (Vendor Managed Inventory) or to the buyer (Buyer Managed Inventory). The process of coordinating collaboration can be readily facilitated by the Internet, and we call this application e-collaboration.

These applications provide the base case for e-Business and a new generation of information systems. A common thread across these diverse applications of supply chain integration is that it supports cross-enterprise coordination in a supply chain, beyond the traditional ERP system. Depending on who controls the information hub, one can classify information hubs into three models – 'e-Market' serving as a marketplace for multiple buyers and multiple suppliers; 'e-Buyer' controlled by a single buyer for multiple suppliers; and 'e-Supplier' controlled by a single supplier for multiple buyers. Alternatively, one can divide information hubs into 'market-centric' (or many-to-many) and 'company-centric' (or one-to-many), depending on whether they are jointly or individually owned.

Now we turn to a full discussion of e-Business that supports supply chain integration over the Internet.

3.1 e-Commerce

Our definition of e-Commerce goes beyond the Business-to-Consumer (B-to-C) interface to include the backend processing of transactions in the supply chain as well. Indeed, the Internet provides a natural setting to link supply chain partners for delivering a product or service in tight coordination. Examples of e-Commerce include Amazon.com, eToys and E*trade. The example of Cisco serves as a case in point.

Cisco Connection Online (CCO) is the largest e-Business site in the world, generating 80% of Cisco's revenue - over $15 billion in 1999. CCO is in essence an information hub that links Cisco with its customers. CCO provides customers with almost everything they need to transact business with Cisco. Through CCO's self-service configuration and order placement system, customers can research pricing, estimate lead times, configure order status, access invoicing and account receivable information, and sign up for service. On the backend side, Cisco streamlined internal operations of order fulfillment in full integration with its front-end order capture by extending its communications to roughly 100 contract manufacturers and suppliers. Once an order is placed through CCO, the backend operations are coordinated via the second hub SupplyWeb. More than 65% of the orders are directly delivered to the customers without Cisco employees ever physically touching them. The Internet-enabled capabilities have reduced order entry cycle time from 1 week to less than 3 days and order-acknowledgment cycle-time from 12 hours to 2 hours. In addition, customer self-service enables Cisco's sales force to focus more on the relationship aspect, rather than the administrative aspect, of its customer relationship. Since implementing CCO, Cisco's customer retention rate is a record 87% and they have experienced a 52% improvement in customer satisfaction. Cisco's estimates of annual savings exceeded $800 million last year.

Another cross-enterprise activity related to e-Commerce is order tracking. Supply chain partners report their order status to the database at the information hub, so that the status is kept current no matter who is handling it at the moment. Several companies market technologies (e.g., Savi Technology) or services (e.g., Descartes) to track and trace orders and resources (like shipping containers) throughout the supply chain. Data are entered into the system by scanners using bar-code or radio-frequency technologies at numerous checkpoints, forwarded to the hub and made available to authorized users. Many logistics service providers like Federal Express and UPS also offer order tracking systems for partners to track and trace orders at all times – be it at a warehouse, at sea, at customs service, or on a truck to the customer. Thus, the platform allows supply chain partners to communicate in case of delays, shrinkage or discrepancies. Another example of an information hub for order tracking is BOG Gases, a distributor of industrial gases (Tedeschi (1998)). The hub is accessible only to select BOG employees, suppliers and customers. Through the hub, BOC's customers place orders to BOG by choosing from an online catalog of gases, and BOG in turn places orders to its suppliers. As the order progresses, suppliers and BOG update the order status. The information hub has cut administrative and inventory costs by avoiding miscommunication, while delivering enhanced service to customers.

Remote sensing, testing and diagnosis are additional examples of e-Commerce activity. Software companies like Norton offer a remote maintenance service on PC products (e.g., tuneup.com). A subscriber of the service would allow the service center to remotely collect data on her computer. The service center (operating as the information hub) will electronically check her computer for computer viruses and terminate them if contamination is detected. They may also advise and help the subscriber to install software upgrades, hardware drivers, and program add-ons specific to her computer. Cisco represents another example of utilizing the hub structure for servicing products. According to the "Autotest" program, Cisco's suppliers run software routines that perform quality tests at their local test cells. The test data are sent over the Internet to Cisco, so Cisco engineers can remotely monitor and control test cells. This enables them to resolve problems that the suppliers themselves cannot diagnose. The standardized test results across the entire supply base allow Cisco to scale the activity rapidly and obtain valuable information about their products that might not be available without such arrangement.

3.2 e-Procurement

e-Procurement is the set of Internet applications by which buyers and sellers find each other and transact according to some pre-specified protocols, and involves private or/and public marketplaces. Since a typical manufacturing company needs to procure thousands of products from hundreds of suppliers, the Internet can help such a company manage the complexity of the procurement process. Numerous companies including Ariba and CommerceOne offer web-based, enterprise procurement solutions that dynamically link the buyer into real-time trading communities over the Internet. They also automate the internal procurement process from requisition to order, as well as the supplier interactions from order to payment. The solutions enable their client companies to reduce operational costs and increase efficiency by automating the entire indirect goods and services supply chain.

In the electronics and high tech (EHT) industry, Converge (formerly known as eHITEX Exchange), e2open and eConnections operate electronic marketplaces for trading direct parts and components. They operate as open e-Markets matching buyers and sellers. By contrast, Digital Market (now part of Agile Software) offers an e-Buyer (named Agile-Buyer) solution for a large buying organization. The solution includes sophisticated data storage and manipulation tools to accomplish complex purchasing tasks – such as part list management, quoting, decision-making, ordering, order change and order confirmation – in hours, instead of days. As the third example, Chemdex (now renamed Ventro) operates a marketplace for biological and chemical reagents for life sciences research needs. Through the e-Market-type hub, pre-qualified buyers gain access to 80% of the leading life science suppliers, along with supporting services including a precision search engine, detailed product information, and online order tracking.

In yet another example, Instill Corporation, a Silicon Valley startup company, employs an e-Buyer model for order capture and processing for the foodservice industry. Its mission is to develop easy-to-use services that lower costs and provide valuable information for all members of the foodservice supply chain, helping to achieve the goals of the Efficient Foodservice Response (EFR) initiative (see CSC, 1997). The company improves upon the traditional time-consuming, error-prone purchasing systems, and helps lower costs for the industry's entire supply chain. Its secure and user-friendly client program allows retail customers (i.e., restaurants and other foodservice operators) to place purchase orders on the web. The orders are forwarded to distributors and manufacturers, according to the rules specified by the retailers. Thus, Instill's server (called Instill Purchase Web) serves as the information hub that links buyers and suppliers in the food service market. The web also offers a purchase tracking service for multi-unit foodservice operators, by allowing operator executives to view up-to-the minute unit purchasing activity for better control. Instill's user-friendly format offers standardized reports to verify contract pricing, track rebates, and monitor unit buying compliance. Further, the manufacturers have access to the aggregate demand and tracking data showing how their products move through each distribution channel – data that were not previously available.

3.3 e-Collaboration

By e-Collaboration we mean the use of the Internet among business partners beyond transactions. Unlike e-Commerce or e-Procurement, whose functions are well defined, e-Collaboration exists in a variety of functions. Examples are different models of information sharing, collaborative decision-making, and product change management.

Information Sharing. An information hub can also serve as an efficient platform to share information among supply chain partners. For example, Baker Street Technologies, a Toronto-based startup, offers a web-based platform that provides a real-time link among supply chain partners. Its information hub technology offers cross-enterprise visibility of supply chain activities. In one implementation, partners share and view purchase orders, sales orders, invoices, checks and other business documents over the Internet. Only the directory and high-level data are kept at the hub, while detailed information and documents are stored at the local sites. When there is a legitimate request, the hub offers the summary or aggregate data. By double-clicking on a data item, one can drill down to the document level and access the local data. In this way an integrated view of supply chain status is collected from disparate information sources and projected on the web site. Similarly, Adaptec, a fab-less semiconductor company also relies on advanced Internet-based solutions to exchange information and coordinate their production plans with their supply chain partners. Using a software called Alliance developed by Extricity (now part of FreeMarket), the company communicates in real time with their foundry TSMC (Taiwan Semiconductor Manufacturing Company) and their assembly partners Amkor, ASAT and Seiko with information such as detailed and complex design drawings, prototype plans, test results, and production and shipment schedules. This arrangement greatly facilitates their ability to be aware of demand and supply levels, and can respond quickly to potential mismatch problems. It also helps to shorten their new product development times. With the use of Alliance, Adaptec's cycle time was cut by more than half.

Collaborative Planning. The Internet provides a system architecture to implement collaborative decision making in a cost-effective way. Several companies (e.g., American Software and Syncra) have developed an information hub that facilitates knowledge sharing and collaborative decision making in the spirit of CPFR, or Collaborative Planning, Forecasting and Replenishment (see Syncra (1998)). Supply chain partners first exchange product forecasts and replenishment plans. Then, its technology synchronizes and develops new agreed-upon plans that closely match supply with market demand. As a result, they can jointly reduce inventory costs and raise customer service levels. Nabisco and Wegmans had successfully implemented a pilot of CPFR, with very encouraging results. The total snack nut category sales went up by 11% while the corresponding sales at other retailers actually declined by 9% in the test period. Nabisco's leading brand Planter's saw its sales increase by 40% as a result of better planned promotions and discounting given to Wegmans stores, which was enabled by the collaborative efforts in replenishment. Finally, Nabisco's warehouse fill rate increased from 93% to 97%, while inventory dropped by 18%. Several other pilots are now under way at Schnuck Markets, Kmart, Circuit City, Procter & Gamble, Kimberly Clark, Sara Lee, and Wal-Mart. Yet another company applies a similar idea to resource planning. Extracting data from multiple partners, their system allows comparative cost analysis on the supply chain level – including tradeoffs between lead time versus fill reliability, and decreased assets versus increased throughput. It also facilitates demand and supply analysis, so the original equipment manufacturer can understand final demand and supply constraints. It additionally analyzes target inventories and safety capacities to meet customer commitments and achieve targeted profitability.

New Product Development. The information hub can also be used to deliver the efficiency and speed demanded in new product development and product change management. Agile Software (www.agilesoft.com), for example, facilitates collaborative product development using the Internet. As product life cycles became shorter and shorter, managing product rollovers is now a routine challenge faced by many high tech companies. Product rollover, defined as the transition from one version of a product to its successor, is often a vulnerable time for a company. One of the major risks in product rollover is the time taken to have all the new parts ready for the rollover. Engineering changes involved in rollovers may require both new suppliers, new bills of materials, and new requirements for existing parts. Agile Software has been able to help companies like Dell, Lucent Technologies, PairGains, WebTV, and Flextronics to use its Internet-based software systems so that engineering changes can be made effortless. Another example of this application is in a major construction project, where numerous geographically-dispersed partners – project teams, architect, auditors, accountants, suppliers and subcontractors – communicate and collaborate. The web site serves as the information hub (using the solution by Framework Technologies) for the project, cutting across geographic and organizational boundaries of a supply chain. Authorized team members can review, contribute and comment in real time on project-related documents of various formats like text documents, CAD files, schedules, spreadsheets, photos, database queries, and other web-enabled applications. The hub maintains the revision history of CAD models and other documents. Cisco also has a similar application. Notifications of changes are automatically sent via e-mail to appropriate team members. Changes in bills of materials are also broadcast to suppliers via the CCO site. According to Cisco, this has reduced engineering change order (ECO) cycle time from 25 days to 10 days within the last 4 years. This improves quality significantly and reduces inventory write-offs, and ultimately, improving overall profitability.

4. Implementation Issues

We review several issues related to the implementation of the information hub.

As mentioned above, the information hub has three models – e-Market, e-Buyer and e-Supplier. Prom the economic welfare point of view, e-Market is likely to yield the maximum efficiency to the community of the supply chain. By sharing a single hub, the supply chain will avoid redundant investment in creating multiple hubs. It also avoids the complications of the e-Buyer (or e-Supplier) model where suppliers (or buyers) have to comply with different rules and standards associated with different hubs. Further, all transacting parties will enjoy the benefits of one-stop shopping. On the other hand, e-Market may be more difficult to implement. Many parties (often competing parties) may have difficulties in reaching an agreement to share the same hub. A common concern is competition among partners over the ownership of customers. A distributor may justifiably worry about the possibility that the hub model will link manufacturers directly to end consumers and ultimately eliminate the role of intermediaries. Even original equipment manufacturers (OEMs) may be vulnerable since distributors may offer products to end consumers in alliance with contract manufacturers, thereby bypassing OEMs. If this concern prevails, e-Buyer or e-Supplier will be easier to implement than e-Market.

The information hub and e-Business can be an extension to ERP systems. Moreover, for a large decentralized company, the information hub can play a hybrid role of internal and external systems. Consider a company again called Acme selling electro-mechanical parts. Acme has a highly decentralized organizational structure – due to its growth by acquisitions. Acme felt that they could further exploit potential synergy in operations. While divisions commonly shared customers and suppliers, their logistics and information systems remained isolated from each other. Four years ago, one division successfully implemented an ERP system originally designed only for its product lines. Over time other divisions with legacy systems and different ERP systems approached them to connect to the ERP system for logistics coordination. Now the ERP system has become a de facto information hub linking multiple internal divisions and external entities like customers, suppliers and logistics partners. Currently, the ERP system handles 50% of the transactions of Acme, but its target is 80% within a year. In this case, the ERP system has evolved to the information hub without initial planning.

The information hub does not necessarily represent a single physical entity. As distributed object-oriented computing is developed, individual functional modules may reside at scattered locations and be invoked only as needs arise. Thus, the information hub could exist only as a logical entity – physically, it is a collection of nodes in the network.

While we focus our discussion on information flows, the idea can be extended to include the logistics hub. A logistics hub is a merge-and-fork point in material flows. A logistics hub subsumes an information hub, but involves the additional dimension of physical material handling. Examples include supplier hubs at manufacturing companies and the integrated logistics hub operated by third party logistics providers.

There are numerous hurdles to information hubs. The first and foremost challenge is that of aligning incentives of different partners. It would be naive to think that supply chain integration will automatically increase each partner's profit. In fact, each partner is wary of the possibility of other partners abusing trust and reaping all or a lion's share of the benefits from supply chain integration. Even when each partner is guaranteed a positive gain in return for participation, partners may haggle over how to split the gain. This may potentially lead to a failure to agree – i.e., resulting in prisoner's dilemma. Thus, trust and cooperation become critical ingredients in a supply chain partnership. On the other hand, trust needs to be rationalized by a relevant economic return.

As mentioned earlier, supply chain integration involves sharing information and knowledge, re-allocation of decision rights, and organizational linkages. When information is shared among supply chain partners, special attention is needed to ensure that the information be kept confidential among authorized personnel, and that the information be used for its original intent only.

Technology is another constraint to the establishment of an information hub. Implementation of a cross-enterprise information system is costly, time-consuming and risky. Partners may not agree on the specifications of the technical system, and how to split the cost of investing in the system.

Finally, companies have to face the challenge of extending or replacing their legacy systems to fit into the new system. Every organization has inertia in committing to a major change. This inertia is well justified in many cases, since the main source of the inertia is the lack of a clear vision and quantitative analysis to direct and guide their e-Business strategy.

5. The Future Trend - It's Intelligence

As various applications of e-Business are implemented, we believe the next trend of e-Business is *intelligence* at the supply chain level. The Internet revolution has made a deluge of data available at managers' terminals. But it remains a challenge to systematically extract meaningful information from the data, present it in a user-friendly interface, and optimize on actionable parameters. Luckily, various performance metrics for the supply chain can be obtained easily as a byproduct of the information hub model. For example, two startups SeeCommerce and Harmony Software provide supply chain partners with supply-chain-wide metrics such as fill-rates, percentage of complete orders, turnaround time, order-level costing, forecast accuracy, customer support service level, and stock levels at each site. Data collection spans multiple applications across enterprises - by geography, product line, and time. This information can be used for benchmarking against the best practices and improving the overall performance of the supply chain. It can also be used as an alert system, so that anytime some key metrics exceed some pre-specified thresholds, alerts are sent to the appropriate entities. Another key benefit of supply-chain-level performance measurement lies in the reduction of printed reports and administrative/distribution costs by making information accessible to authorized personnel via a web browser.

Optimization as the next step of intelligence is already incorporated in some software engines by companies like i2 and Manugistics. Freightwise is an example in this direction. It operates an open e-market for transportation services matching buyers and sellers. But Freightwise provides other value-added services including intelligence. It allows a shipper to view various options arising from the bids at the exchange and makes a recommendation regarding which options to take to minimize the cost under a set of constraints. Another example is Cisco, which, as described earlier, has used the Internet effectively as a means for procurement and sales management. But the company wants to do more. The newest development at Cisco is the new eHub that contains intelligence. The new eHub, to be built by Manugistics, is supposed to be smart enough to detect problems (such as supply shortage, schedule gone out of control, and other disruptions in the supply chain), identify the right parties, and create new plans of actions.

If this trend continues, the decision environment in the manufacturing sector will look more like that in the financial sector. In the past, manufacturing has operated on monthly or weekly cycles. MRP was run monthly, and the information feedback (like consumer market trends or material price movement) came a week or a month later. While several decisions were made minute by minute on the plant floor, decision makers were poorly equipped with relevant information. Conditions beyond the manufacturing floor or the company boundary were taken as a black box, so in the absence of data, optimization on a global scale was limited or prohibited. By contrast, a financial trader working in the foreign currency exchange market, for example, operates on the clockspeed of seconds to exploit fleeting opportunities of arbitrage, armed with real-time data feeds and dynamic analysis tools. Now that the Internet creates a new environment for manufacturing with real-time market data and various operational instruments (like multiple secondary markets, long-term contracts, forwards and futures), manufacturing will also require similar information systems to support such dynamic decisions and actions. It is our belief that the next generation of e-Business products and services will address these needs.

6. Conclusion

This article has studied various e-Business applications in three classes (e-Commerce, e-Procurement and e-Collaboration) under the new information hub architecture, and discussed the potential of e-Business towards intelligence and optimization.

In retrospect, the first generation of e-Business applications has improved the supply chain efficiency through the ability to access extensive information, automate workflows, open a direct channel, and link multiple suppliers and customers. Such applications, in the areas of e-Commerce, e-Procurement and e-Collaboration, have reduced the operating costs, shortened lead-time, improved customer service, and eliminate transaction errors in a supply chain. As such, significant values have been achieved.

On the other hand, e-Business is supposed to realize much greater benefits and values than what we have seen so far. The early ventures of e-Business (mostly in the e-Commerce area) have achieved some limited success, but not up to our original expectation. Pure content access and transaction efficiency are not the greatest payoff that could be obtained with e-Business. Instead, the opportunity should lie in the ability to optimally synchronize and coordinate the supply chain, to collaborate among supply chain partners for faster product development and introduction, and to create new services to penetrate new markets.

Despite these disappointments, the next few years will see an explosion of e-Business applications as visionary companies develop new paradigms for the future. Such visionary companies as Cisco and numerous startups have already demonstrated ample opportunities in e-Business. It may be only a matter of time that more companies will soon follow their lead into the information age.

Acknowledgments

The authors would like to thank the two reviewers for numerous comments on the earlier version.

References

Bock, G.E., 1998. "Coordinating the Web of Inter-Enterprise Relationships: The Business Benefits of Real-Time Customer Service Solutions form Service Track," Patricia Seybold Group, January 1998.

Computer Sciences Corporation (CSC), 1997. *Efficient Foodservice Response: Enabling Profitable Growth in the Food-Prepared-Away-from-Home Industries,* Food Distributors International - NAWGA/IFDA, Inc., et al.

Kurt Salmon Associates Inc., 1993. "Efficient Consumer Response: Enhancing Consumer Value in the Grocery Industry," Food Marketing Institute, Washington DC.

Lee, H.L., 2000. "Creating Value Through Supply Chain Integration," Supply Chain Management Review, 4, September/October, 2000, 30-36.

Lee, H.L. V. Padmanabhan and S. Whang, 1997. "The Bullwhip Effect in Supply Chains," *Sloan Management Review,* 38(3), Spring 1997.

Lee, H.L. and S. Whang, 2000. "Information Sharing in a Supply Chain," *International Journal of Technology Management,* Vol. 20 (3/4).

Syncra Software Inc, 1998. "Collaborative Planning, Forecasting," & Replenishment (CPFR)," One Cambridge Center, Cambridge MA 02124.

Tedeschi, B., 1999. "The Net's Real Business Happens .Com to .Com," *The New York Times on the Web,* April 9, 1999.

Chapter 2

CUSTOMER SERVICE MODELS FOR BRICKS, CLICKS AND IN BETWEEN

Douglas J. Thomas
Smeal College of Business Administration
The Pennsylvania State University
University Park, PA 16802
dthomas@psu.edu

1. Introduction

The Internet has sparked substantial growth in consumer-direct delivery. While the phenomenal growth of "pure play" Internet retailers such as Amazon.com certainly strikes fear in the hearts of traditional Bricks-and-Mortar retailers, these retailers can still offer value to consumers through better service, including instant gratification, easier returns/exchanges and permitting the customer to physically examine the product. A growing number of established retail companies are choosing to augment their physical presence with an online presence, a "bricks and clicks" strategy, in an attempt to capture the best of both worlds. Mall standards such as The Gap and J.C. Penney have had some success employing such a strategy (see Lee (1999)).

Of critical importance to firms using any or all of these delivery channels is outstanding customer service. A recent customer service study from Sybase[1] asserts that a dissatisfied customer will tell eight to ten people about his or her experience, it costs six times more to sell to a new customer than to sell to an existing one, and seventy percent of complaining customers will do business with the company again if it quickly takes care of a service failure.

Any inventory model with stochastic demand must capture the possibility that demand may not be completely satisfied from available in-

[1]`www.sybase.com`

J. Geunes et al. (eds.),
Supply Chain Management: Models, Applications, and Research Directions, 19–32.

ventory and that the seller of the good bears some cost of not satisfying demand. Such models seek to determine appropriate stocking and replenishment behavior, generally trading off expected stockout costs with expected inventory carrying costs. While throughout this document, we refer to selected work from the stochastic inventory literature, it is not our intention to rigorously review this area. Silver et al. (1998) offers an excellent treatment of basic inventory models.

As the product and service marketing experience becomes more personalized, better information about individual consumers is available. Customer information that may be available includes demographic information, previous buying behavior as well as the consumer's hobbies and interests. Fully incorporating this information may include customizing product and service offerings.

Our purpose in this paper is not to discuss all the complexities of individual customization, but rather to examine how the presence of multiple delivery channels affects both the likelihood and cost of certain kinds of service failures. In some environments, this will simply affect the estimation of stockout costs. In others, traditional stochastic inventory models with a single stockout cost will be inadequate. In section 2, we present a taxonomy of consumer-facing service failures, considering four kinds of stockouts and four sales channels. We discuss the manner in which these sixteen "stockout types" vary in both (per unit) cost and probability of occurrence in section 3. In section 4, we discuss environments where modeling this level of stockout detail might be useful.

2. Stockout Types

We focus our discussion on single item systems where customer service is measured by a *per unit* stockout cost. Many authors have pointed out that there is an equivalence between stockout costs and a service level constraint (see McClain et al. (1992) or Nahmias (1989)). It is worth remembering this fact since many managers seem quite comfortable with specifying a service level, but not specifying a stockout cost. Since there is an implied per unit cost associated with a fill rate constraint, the issues raised apply to these constrained models as well. The concepts discussed also apply to systems where multiple items are considered, such as *order* fill rate models (see Song (1998)) and models where the service penalty cost is incurred per event rather than per unit.

Chen and Zheng (1993) and Cetinkaya and Parlar (1998) examine models with more general backorder costs (not linear) and extend some results to quasi-concave holding and backorder costs. Also, there is significant work that addresses a mix of lost sales and backordered demand (see Moon and Choi (1998))

We consider four kinds of stockouts:

1 **Lost Sale (LS)** the desired good is not available, and the buyer leaves the system unsatisfied;

2 **Backorder (BO)** the desired good is not available, and the buyer agrees to wait and complete the sale when the good becomes available;

3 **Buyer Substitution (BS)** the desired good is not available, and the buyer selects an alternate product;

4 **Seller Substitution (SS)** the desired good is not available, and the seller offers comparable or superior product at the price of the originally desired good.

Simultaneously, we consider four retail selling channels:

1 **Bricks only (BRK)** the seller offers the product only in physical stores;

2 **Bricks and clicks, store order (BCS)** the seller offers the product both in physical stores and online, and the buyer goes to the store to obtain the product;

3 **Bricks and Clicks, online order (BCO)** the seller offers the product both in physical stores and online, and the buyer goes online to obtain (order) the product;

4 **Clicks only (CLK)** the seller only offers the product online.

We only consider systems where either buyer substitution or seller substitution occurs, not both. The justification for this is that if seller substitution is offered, no buyer substitution will take place since the seller substitution involves offering a superior product. It is possible that there are systems where both kinds of substitution are possible. The simplest such case is where the buyer in a physical store does not find the desired product on the shelf and selects another product without speaking with the sales staff (who would have made the buyer a better offer; see Hsu and Bassok (1999) and Bassok et al. (1999) for discussions of seller substitution.)

A buyer will choose to go to a store for several possible reasons. At the store he can physically touch, feel, test and operate the product. He may enjoy the interaction and sensory experience of the store (some people actually enjoy going to the mall!). He may have security concerns regarding Internet shopping. He may want the product instantly, and he may favor the post-sale customer service offered by physical stores. In fact, in a recent study by Jupiter Communications, 37% of consumers stated that they would be more likely to shop online if returning product was easier. Conversely, a shopper will choose an Internet channel for potentially lower cost, ease of comparison shopping (it is easier to visit multiple selling locations) and convenience of home delivery. Interestingly, a recent empirical study of online buying in the hospitality industry by Shankar et al. (1999) suggests that while price search time increases in the online medium, non-price information available via the Internet actually leads to less price sensitivity. That is, customers may shop around more, but they tend to be even more brand loyal.

While the seller has control over the inventory stocking decision and the choice of channels, she has only limited control over the kind of stockout that will occur. That is, she can offer to make it easy for the good to be backordered, perhaps offering expedited delivery or another incentive, or she could offer equivalent or superior substitute products. Still, it is the buyer that ultimately determines the kind of stockout that occurs. In the next section, we discuss how the different sales channels affect both the cost of service failures and the likelihood of certain buyer actions.

3. Impact of Channels on Stockout Cost and Likelihood

We use the generic term *stockout* to represent any unsatisfied demand at the time the demand occurs. For each stockout, there are costs that are relatively easy to quantify such as lost marginal profit, cost to expedite an order or higher cost of a substituted product, as well as costs that are difficult to quantify that capture the effect of poor service on the future buying behavior of the dissatisfied customer and the general reputation of the seller. The latter of these are known as *goodwill* costs. The combination of these direct costs and goodwill costs gives a penalty cost associated with unsatisfied demand.

Many authors have discussed the difficulties in estimating goodwill costs. Recently, Ishii and Konno (1998) address this problem by developing a fuzzy newsvendor model, where the stockout cost is a fuzzy number. Jones (1999) addresses the problem of estimating lost sales due to balking. That is, if a customer does not see the desired product, they may leave the system and the seller may not know that a lost sale occurred.

While goodwill costs are difficult to estimate, it is important to note these costs should not necessarily be proportional to the magnitude of customer dissatisfaction. The goodwill cost is an attempt to quantify something that, were we omnipotent, could be quantified. The seller is concerned with losing future sales, both from the dissatisfied buyer and from people that buyer might tell about his experience. If a seller has a strong competitive advantage (price, quality, etc.), a buyer can be very unhappy about the service he receives from that seller, but he will still return to that seller. In a more competitive environment, a slightly dissatisfied buyer could switch to another seller forever. It is the level of dissatisfaction combined with the competitive environment that affect goodwill costs. Consider the simple example described in Table 2.1 with a single selling channel, perhaps a retail store, and three kinds of customers, very loyal, somewhat loyal and fickle.

Customer Type	Behavior	Cost
Very Loyal	Customer will backorder the desired good	No lost sales. Just the cost of recording and satisfying a backorder
Somewhat Loyal	Customer will buy elsewhere, but come back in the future.	One lost sale.
Fickle	Customer will buy elsewhere and not return.	Many lost sales; this sale and future potential sales both from this customer and potentially his friends

Table 2.1. Estimating Stockout Costs for Different Customers

If a retailer can estimate the distribution of customer types, she can compute an average stockout cost and apply basic stochastic inventory models to determine appropriate safety stock levels and reordering policies. If information about each customer is available, the seller might ration stock or make different substitution offers based on the customer profile. For example, if stock is expected to be in short supply, the retailer might choose to not satisfy a customer demand if she knows that customer is very loyal. We address this and other related issues below in section 4.

3.1 General Cost Issues

A primary difference between lost sales and other stockout types is that when sales are lost, the marginal profit is lost. (For the purposes of our discussion, we assume that the seller is not a new e-tailer, selling each unit at a loss, trying to make up the difference in volume.) For any channel, lost sales should always be the most expensive kind of stockout. If not, whatever kind of stockout was more expensive should never be performed (the seller should not backlog or substitute). This does not mean this cannot happen. Recall that when we say stockout cost, we mean to include both direct and goodwill costs. If the goodwill costs have been overestimated, a seller will do more to satisfy the buyer than she should.

Other than those distinctions, it is rarely possible to provide an ordering of different types of stockout costs, based on their relative magnitudes. With all else equal, the goodwill cost for buyer substitution is higher than the goodwill cost for seller substitution since the customer will appreciate the special attention associated with seller substitution. Of course, while goodwill cost will be lower with seller substitution than with buyer substitution, there will be higher product related costs with seller substitution.

3.2 Lost Sales Across Channels

Lost sales costs for either online channel (BCO or CLK) are quite high. The Sybase study mentioned previously asserts that the probability of selling a product to a new customer is 15 percent, whereas the probability of selling to an existing customer is 50 percent. Lost sales costs for online orders are high relative to store lost sales costs since the store will have a location advantage. That is, a customer may be dissatisfied with the store inventory shortage, but there is a reasonable chance he will return since there may be limited choice in his geographic area. BCS lost sales costs are slightly lower than BRK since the bricks and clicks seller offers greater channel choice to the consumer. The probability of a BRK lost sale is slightly lower than with BCS since it may be easier for the buyer to place a backorder. (The retailer may have Internet kiosks in the store.)

The environment is more competitive for online sellers since they have no location advantage. For both BCO and CLK, the lost sales cost is high since a dissatisfied customer may never return. Similar to the physical store situation though, lost sales costs for BCO may be lower than for CLK since the buyer may have chosen the seller because of the seller's physical presence. In this case, the seller is more likely to retain future business from that customer. Given these observations, we hypothesize the following orderings for lost sales cost and probability:

$$\text{Lost sales cost: } CLK > BCO > BRK > BCS \tag{2.1}$$

$$\text{Probability of a lost sale: } CLK > BCO > BRK > BCS \tag{2.2}$$

3.3 Backorders Across Channels

In all cases, the magnitude of backorder costs and the probability of backorders occurring depend on the duration of the customer wait. Backorder costs for BRK orders range from moderate to high, depending on the importance of instant gratification. The probability of a backorder is probably lowest with this channel since the customers are more likely to want the product immediately.

For BCS, both the cost and probability of a backorder are lower than for BRK, since it is easier for the customer to place a backorder (assuming the customer can place an online order from the store.) For both BCO and CLK, the customer was going to wait anyway, so (again depending on the length of the wait) the customer may not be that dissatisfied with a backlog situation for an online order. For this stockout type, we hypothesize the following cost and probability orderings:

$$\text{Backorder cost: } BRK > BCO > BCS > CLK \tag{2.3}$$

$$\text{Probability of a backorder: } CLK > BCO > BCS > BRK \tag{2.4}$$

3.4 Buyer Substitution Across Channels

Stockout costs associated with buyer substitution for BRK can go from almost zero to very high. If the buyer chooses a substitute product, the marginal profit is probably similar; thus the stockout cost depends primarily on the goodwill cost. Consider a wine store. A buyer on his way to a dinner party may enter a wine store with a particular bottle in mind. If that bottle is not available, many substitutes are available. The buyer is probably not too unhappy, and he will likely return in the future. Conversely, a buyer might be unwilling to go to another seller given his time constraints, but he will be quite dissatisfied and may not return to that seller, particularly if there is substantial competition. A gas station that has run out of regular fuel and only has premium is one example.

For BCS orders, buyer substitution stockout costs are similar to BRK stockout costs. The primary difference between BRK and BCS is that buyer substitution is less likely for BCS since it is relatively easy to backorder the desired item by placing an online order from the store.

For BCO orders, buyer substitution stockout costs are low compared to store channels. The reason for this is that the buyer is unlikely to feel forced into a substitute product he does not want. If no acceptable product is available, a lost sale is very likely. This means that while the buyer substitution stockout cost is low, the probability of such substitution is also low. Instead, the shortage will frequently result in a backorder or lost sale. Buyer substitution stockout costs for CLK orders are similar to those for BCO. Again, we hypothesize cost and probability orderings:

$$\text{Buyer substitution cost: } BRK, BCS > BCO, CLK \tag{2.5}$$

$$\text{Probability of buyer substitution: } BRK, BCS > BCO, CLK \tag{2.6}$$

3.5 Seller Substitution Across Channels

In nearly all cases, the customer will be happier with seller substitution than with backordering, implying that the *goodwill* cost is lower for seller substitution. However, some product-related cost must be absorbed by the seller; typically a more expensive product is sold for the same price. Since goodwill costs will be low in this case, the cost of the alternate offering will strongly influence the total stockout cost. Seller substitution is most likely with BRK since the buyer may be seeking instant gratification. The probability of seller substitution with BCS is lower since the buyer can easily backorder the good. Since the cost of seller substitution depends on the alternate offering, we hypothesize a cost ordering only:

$$\text{Probability of seller substitution: } BR > BCS > BCO, CLK \tag{2.7}$$

Table 2.2 summarizes our discussion of the differences in both cost and probability of the different kinds of stockouts.

4. Customer Service Model Implications

4.1 Inventory "Coverage" for a Bricks and Clicks retailer

A bricks and clicks retailer would have an existing network of stores as well as extensive online operations. Given the different activities required in bricks and clicks, it is not uncommon for the logistics networks for the two operations to be quite independent. Consider the the case of BarnesandNoble.com, where the online operation was spun off as a separate corporation. (It is interesting to note that Barnes and Noble and bn.com have grown closer since the spin-off, with the physical stores now accepting returns and exchanges from online purchases.) In fact, there

	Bricks	Bricks and Clicks - Store	Bricks and Clicks - Online	Clicks
Lost Sales	Expensive but not as bad as lost sales for online channels since the store may have a geographic advantage.	Nearly as expensive as BRK, but there is a better chance the buyer will accept a backorder (shipped from online facility)	Very expensive since customer is a click away from another site. Not as costly as clicks only lost sale since there is some value to the consumer of doing business with B/C company	Very expensive. Buyer is a click away from another e-tailer and may never return
Backorder	Buyer is unhappy. They may have come to the store for instant gratification	Not as bad as BRK since the order can be placed through the online mechanism	If the wait is not long, the buyer is not too unhappy. He was expecting to wait anyway	Same as BCO
Buyer Substitution	The probability of buyer substitution is highest with BRK. Stockout costs can be very low (wine store) or very high (gas station).	Probability of buyer substitution is slightly lower for BCS than BRK since the buyer may be able to get the desired product online (a backorder).	Probability of substitution is lower than either store channel since if the buyer must wait, he may choose to wait for the desired product. Stockout cost is low since if the buyer really did not want to substitute, he would click elsewhere.	Similar to BCO
Seller Substitution	For all channels, stockout cost depends on the quality of the alternate offering. It is probably the case that substitution is more likely (with equivalent offerings) for store (versus online) channels			

Table 2.2. Sixteen Stockout Types

is a continuum of bricks and clicks operations. Gateway country retail outlets are "service centers" where *all* products are custom-configured and delivered to the home. The purpose of the retail outlet then is to provide information, let the potential consumer examine similar products and provide post-sales service. At the other extreme, it is unlikely that a Wal-Mart shopper facing an out of stock product will backorder it by placing an online order from the store.

Now, consider a bricks and clicks operation somewhere between these extremes, such as The Gap. A customer would go to such a clothing store (rather than buying online) to examine the products, try them on for fit, see how the color looks, etc. He may also want the product for immediate use. For a particular style of casual cotton pants (e.g. flat-front, relaxed fit khakis) there are three product dimensions, waist size, inseam and color, leading to many different stock keeping units (SKU). In the absence of online operations, a store manager might simply forecast demand for each item and trade off holding and stockout costs. The problem with that approach is that the number of items is quite large. In fact, to avoid excessive holding costs, Gap stores have a limited selection, with a wider selection of sizes available online. That is, you can buy 32″, 34″,... waist size in the store, but 33″ waist size is available online only.

Clearly there is a demand pooling benefit to selling certain items online only. The risk is that store-going customers wanting in-between sizes may be dissatisfied. One way to address this problem is to use a concept we term inventory *coverage.* Recall that the store-going customer wants three things: to establish fit, see how the color looks on him and get the pants immediately. As long as the store maintained inventory in each waist-color combination and in each waist-length combination, two of the three customer objectives are satisfied. In fact, in the case of pants, we could relax this slightly. That is, for establishing the color appearance objective it may be sufficient to be *close* in size (for example, correct waist size and within an inch or two of the correct inseam.)

This coverage concept moves the store closer to being a service center. Still, the store-going customer typically wants immediate satisfaction, so a retailer using a this size-color coverage concept might want to estimate two stockout costs, the normal item stockout cost and the cost of a coverage stockout. A normal stockout would occur when the customer identifies the desired good, but that good is not available. A coverage stockout would occur when there is insufficient inventory for the customer to find the correct fit or examine the full range of colors. Presumably the cost of a coverage stockout would be greater than the cost for an item stockout, suggesting a two-tiered service level model.

Furthermore, while this store experience is going on, there is an inventory stocking decision for the online distribution system that must be made. If the store commits to supporting such a coverage model, the supporting online operation should have a very high service level. If a customer comes to the store expecting to try on clothes and then place an order online, he will be quite unhappy if he encounters an *online* stockout.

4.2 Using Customer Information

The discussion in the previous section about inventory coverage raises an interesting point regarding customer location. It is conceivable that an online retailer (CLK or BCO) would have information about the location of the online shopper. A bricks and clicks retailer would like to know if the shopper is close to one of her retail locations, or a competitors location. The scenario leading to the highest stockout cost is probably where the shopper is in an area where the seller has no store, but a competitor does have a store. Any service failure could very likely result in that shopper defecting to the retailer with a local presence. (In fact, the author admits to defecting in exactly this manner due to a service failure.) It might be worthwhile for the retailer to make substitution offers to customers in this situation where she has a location disadvantage, but make no substitution offer where she is not disadvantaged.

Recently, Cattani and Souza (2000) study a model where an online retailer reserves stock. It is possible that a retailer could ration stock in a physical store, although this would certainly be more difficult since the product is stored where the customer can see the inventory. Clearly, if a customer is denied inventory they can see, they will not be happy. Certainly an online seller can mask stock information from customers. In fact, consider results from the Accenture online holiday shopping study shown in Table 2.3.

Company Type	Year	Status Provided	In-stock (if status provided)
E-tailer	1999	46%	98%
	2000	36%	89%
Catalog merchant	1999	43%	84%
	2000	41%	90%
Traditional retailer	1999	43%	96%
	2000	36%	91%
Total	1999	45%	96%
	2000	37%	90%

Table 2.3. From the Andersen Consulting/Accenture e-Santa study of 1999 and 2000 Holiday shopping

The fraction of online vendors posting inventory information came down substantially from 1999 to 2000. It seems unlikely that the primary reason for this drop is that retailers are rationing stock based on customer profiles, however it is an interesting decision as to whether the retailer should post this information freely or only when the "add to cart" button is clicked.

5. Remarks and Conclusions

There are many relevant customer service issues we have not addressed of course. The discussion of multiple channels certainly suggests the possibility of jointly determining stock levels for retail outlets and fulfillment centers as well as setting a different price for each channel. It may well be the case that pricing policies that encourage store-goers to take home delivery may improve system performance by reducing system inventory levels. For this to work, savings from this inventory reduction must offset the higher transportation costs associated with home delivery.

The online medium lends itself to promotional marketing, such as special email offers. Letting the system state play a significant role in determining the timing and nature of promotions could lead to significant operational improvement. As a simple example, consider an online fulfillment operation. Sending an email blitz to generate a surge in orders when the facility is working at or near capacity could lead to customer delays, while special offers during a period of underutilization might improve profitability. There is a need for models that tie such email promotions to both inventory availability and system capacity.

Recent discussions with several e-tailers and logistics providers led to a somewhat startling discovery regarding customer expectations. The approximate breakdown of choice of shipping options (for moderate value goods) was 75%, 15%, 10% for ground, two-day and overnight. Many e-tailers have some form of fulfillment time guarantee, such as "in-stock items shipped within 24 hours," however, the focus of the fulfillment center operations is typically on overall productivity (for example lines picked/hour). Often, orders are not prioritized by customer type or shipping mode. That is, when a consumer selects ground shipping, they indicate that delivery time is not critical. Of course a retailer cannot make them wait indefinitely, but their order can certainly wait a day. A consumer choosing one or two-day service has clearly indicated his time sensitivity. If the above 75%, 15%, 10% estimate is close, no one or two day order should ever wait more than a day (assuming inventory is available). Furthermore, there is an opportunity to affect the consumer's choice of shipping mode via price changes or promotions to maximize revenue while maintaining excellent customer service.

We have suggested a taxonomy of customer-facing stockouts in the presence of traditional and online delivery channels. The sixteen stockout types discussed here differ in both cost to the seller and probability of occurrence. As more individual customer information becomes available, more sophisticated models will be needed to help retailers make intelligent stocking and selling decisions. Both practitioners and researchers should be aware of the implications of these different kinds of stockouts. Simple order-up-to and reorder point systems may not be adequate to describe optimal stocking behavior when the role of the store is both a stocking location and a service/product demonstration center.

References

Bassok, Y., Anupindi, R., and Akella, R. (1999). Single-period multiproduct inventory models with substitution. *Operations Research,* 47(4):632–642.

Cattani, K. and Souza, G. (2000). Inventory nuances for e-commerce companies. *Working Paper - Kenan-Flagler Business School, UNC Chapel Hill.*

Cetinkaya, S. and Parlar, M. (1998). Optimal myopic policy for a stochastic inventory problem with fixed and proportional backorder costs. *European Journal of Operational Research,* 110(1):20–41.

Chen, F. and Zheng, Y.-S. (1993). Inventory models with general backorder costs. *European Journal Of Operational Research,* 65(2):175–186.

Hsu, A. and Bassok, Y. (1999). Random yield and random demand in a production system with downward substitution. *Operations Research,* 47(2):277–290.

Ishii, H. and Konno, T. (1998). A stochastic inventory problem with fuzzy shortage cost. *European Journal of Operational Research,* 106(1): 90–94.

Jones, L. K. (1999). Inferring balking behavior from transactional data. *Operations Research,* 47(5):778–784.

Lee, L. (October 18, 1999). Clicks and mortar at Gap.com. *Business Week.*

McClain, J. O., Thomas, L. J., and Mazzola, J. P. (1992). *Operations Management: Production of Goods and Services.* Prentice Hall, third edition.

Moon, I. and Choi, S. (1998). A note on lead time and distributional assumptions in continuous review inventory models. *Computers & Operations Research,* 25(11):1007–1012.

Nahmias, S. (1989). *Production and Operations Analysis.* Irwin.

Shankar, V., Rangaswamy, A., and Pusateri, M. (1999). The online medium and customer price sensitivity. Working Paper 04-1999, eBusiness Research Center, Pennsylvania State University.

Silver, E., Pyke, D. F., and Peterson, R. (1998). *Inventory Management and Production Planning and Scheduling.* John Wiley & Sons, third edition.

Song, J.-S. (1998). On the order fill rate in a multi-item, base-stock inventory system. *Operations Research,* 46(6):831–845.

Chapter 3

B2B MARKETS: PROCUREMENT AND SUPPLIER RISK MANAGEMENT IN E-BUSINESS

Ram Akella

Department of Industrial Engineering/School of Management
and Center for Excellence in Global Enterprise Management
State University of New York at Buffalo
Buffalo, NY 14260
akella@eng.buffalo.edu

Victor F. Araman

Department of Management Science and Engineering
Stanford Unversity
Stanford, CA 94305
varaman@stanford.edu

Jochen Kleinknecht

Department of Management Science and Engineering
Stanford Unversity
Stanford, CA 94305
jochenk@stanford.edu

Abstract This chapter clarifies some of the different types of efficiencies that can be achieved in B2B electronic commerce. We focus on the risk management aspects of these efficiencies, especially those pertaining to supplier or supply chain management related costs. This novel approach incorporates risk management by combining the traditional collaborative framework (of reservation capacity and associated pricing) with a dynamic spot market, both of which are Web-based. By way of contrast, in prior research and industrial use, these two supply modes of collaborative commerce and spot market, have been used separately, in a

J. Geunes et al. (eds.),
Supply Chain Management: Models, Applications, and Research Directions, 33–66.

non-integrated way. We develop a framework that draws upon our research, as well as that of others, suitably adapted by us to this context and interpreted to provide insights into the economics of B2B commerce against this backdrop. We also point out open questions and identify new research directions.

Scope and Definition of B2B E-Commerce

For this chapter, we adopt the following definition presented in Durlacher (2000) and define business-to-business e-commerce as well as business-to-business electronic markets as follows:

Definition 3.1 *Business-to-business e-commerce is commerce conducted between businesses over an intranet, extranet or Internet (i.e., IP networks). This trade may be conducted between a business and its supply chain as well as between a business and other businesses' customers.*

Definition 3.2 *Business-to-business markets are platforms on which B2B e-commerce may be conducted directly between buyer and seller or through a third party.*

We start out this chapter with a brief discussion of several key sources of efficiency in B2B e-commerce, out of which we will mainly focus on risk management approaches. We then supply a framework for B2B e-commerce, and argue that traditional supply chain literature is not sufficient to provide insights into all aspects of B2B e-commerce. We continue with a description of the characteristics of Internet-based marketplaces and present the key results of recent work that fill some of the gaps between B2B e-commerce and the traditional supply chain research. We conclude the chapter with a brief summary of the main results, and sketch on-going research and an outlook on future work.

1. B2B E-Commerce: Sources of Efficiency

It is well known that B2B e-commerce achieves efficiencies through several of the following key mechanisms, as well as others:

1. Process Efficiencies

The standardization of software and formats for exchanging business information creates significant efficiencies, that are exploited in e-business. These include supplier discovery, i.e., analyzing and negotiating purchases with suppliers, price discovery, as well as automating labor-intensive procurement and sourcing processes. Moreover, the establishment of common standards within B2B markets will let companies juggle their suppliers, depending on available supplies.

2. Web Efficiencies

The Internet not only offers hardware and communication cost reduction but also allows for a broader reach, in comparison with EDI (Electronic Data Interchange) services, which provide only point-to-point connections. This is especially important in the supply chain area of B2B commerce, which is traditionally messy, paperwork-intensive and prone to miscommunication. Information visibility, including information required to mitigate the bullwhip effect (referring to the increasing variability of orders as we move upstream in the supply chain; see Lee et al. (1997) for more details), can be maximized through the use of Internet-based systems while maintaining a tight access control for this highly sensitive data. The Internet also opens up the possibility for businesses to sell more of their products across all of the goods lifecycle (raw material, finished goods, second hand goods, scrap, etc.).

3. Demand/Forward Aggregation

Significant procurement cost reductions also result from volume discounts achieving by pooling demands, through enhanced bargaining power or concave cost structures. Obviously, forward aggregation yields similar benefits for groups of suppliers, who cooperate to increase their negotiating power.

4. Value Added Services

Value added services address issues of what companies should buy, who they should buy it from, what price they should pay, when they should buy and where. For instance, trading exchanges are likely to extend their offerings to include settlement and fulfillment capabilities. More explicitly, pre-qualification of quality of suppliers, logistics providers, etc. can significantly lower transaction and supply chain costs and thus create overall efficiencies.

5. Information Aggregation

The Internet is an increasingly global network allowing businesses to reach customers and suppliers in new areas. The ability to combine multiple sources at one portal, as well as web-based search capabilities enable significant reduction in locating suppliers and conducting due diligence, i.e., verifying price, quality, and service characteristics (page 395 in Chopra and Meindl, 2000).

6. Information Sharing

Historically, only limited information was shared between companies. B2B e-commerce technologies, however, allow for real-time communication and data sharing by integrating ERP (Enterprise Resource Planning) and other systems with those of an organization's suppliers and customers thus eliminating duplication and achieving closer matching of demand and supply. At the same time, supply contracts embedded in ERP collaborative commerce software modules increase transparency between cost and profit centers.

7. Price Discovery Mechanisms

Price discovery mechanisms used in B2B e-commerce such as auctions and reverse auctions reduce bargaining and coordination costs, which historically have represented the most significant part of transaction costs.

8. Risk Management

By using the B2B exchange ability to access multiple types of suppliers such as long-term capacity suppliers, as well as spot market suppliers with varying price, quantity, quality, and service characteristics, buyers can become more flexible and are able to manage their supply according to actual demand instead of uncertain forecasts. Likewise, suppliers can now sell to multiple buyers at a given time and have a platform to sell off any leftover inventory or excess goods. These benefits stem from enhanced efficiency in capacity management, through effective supply and demand risk pooling.

In this chapter, we focus on the last four aspects described above, while putting particular emphasis on the last two, to achieve efficiencies and enhanced profits. We next describe a B2B e-commerce framework that helps capture these elements.

2. Framework for B2B E-Commerce

There are many frameworks that have been developed for classifying B2B e-commerce, to describe the differences in policies and functioning of the various forms of B2B hubs. The intent of our specific framework is to include all forms of B2B e-commerce and to indicate the primary types of markets (e.g. components, capacities), where our risk management approaches provide significant economic gains, and possible adaptations to other types of markets. Our approach is a major modification of a framework recently suggested by Kaplan and Sawhney (2000), where our

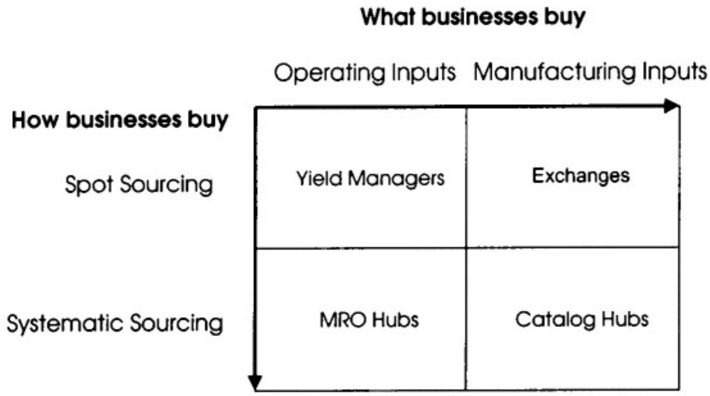


Figure 3.1. B2B Matrix by Kaplan and Sawhney

framework enables the use of existing risk management models, or the identification of gaps to develop new models.

The Kaplan-Sawhney model described in Figure 3.1 above, has the classic 2×2 matrix, with systematic (collaborative) source vs. spot market on one axis, and operating inputs (meaning MRO - maintenance, repair, and operating - resources across horizontal industries, such as computers) vs. specific manufacturing inputs on the other. They classify B2B marketplaces, which they call electronic hubs, or e-hubs, in one of the four boxes of the 2X2 matrix: spot markets for common operating resources like labor, advertising, and manufacturing capacity are called yield managers, whereas on-line exchanges are defined to trade commodity like production inputs, such as steel and energy. So-called MRO hubs streamline the sourcing of low-value goods with relatively high transaction costs and catalog hubs automate the procurement of non-commodity, industry specific manufacturing inputs.

Observe that they categorize capacity under operating inputs, which is somewhat awkward. We will soon observe that our proposed framework is more consistent from a modeling and risk management perspective.

Figure 3.2 describes our framework for B2B e-commerce. Our objective is to classify the different types of B2B e-commerce interactions and markets, and to model the economic gains that a combination of markets can help us achieve, rather than by using one type of market alone.

Observe that while our vertical classifications are the same as in Kaplan and Sawhney in terms of spot markets and systematic sourcing to describe buyer-seller relationships, our horizontal classification is now based on standardized and customized types of goods or service. The reason for this is that we are interested in exploring the economic de-

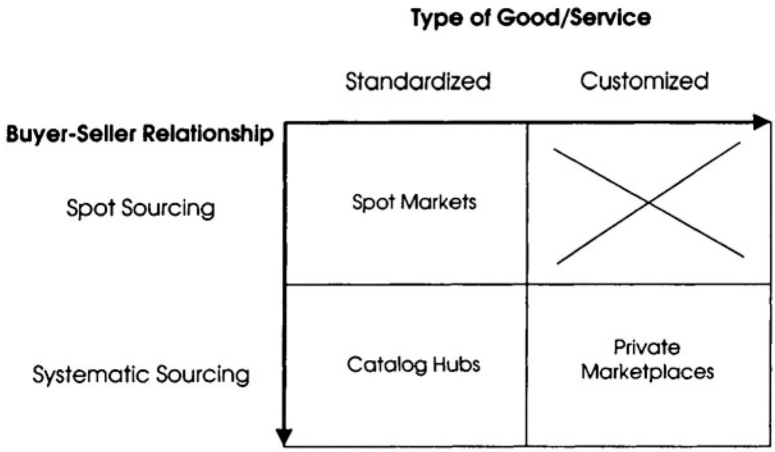


Figure 3.2. B2B E-Commerce Framework

sirability (through risk management) of using possibly more than one market type simultaneously.

We now discuss the specific elements of Figure 3.2.

2.1 Private Marketplaces

These correspond to systematic sourcing, which may be based on long-term contracts, implemented via the Web (or manually), and represent customized products. There are several elements of economics that could potentially preclude web based search and comparisons of customized products from multiple sources:

- Shipment cost or difficulty
- Uniquely engineered products with attendant difficulty or costs if procured from more than one supplier site.
- The award to a supplier of preferred status and long-term relationship, with attendant responsibilities of design and engineering work

Under these conditions, when the nature of the good or service in question is such that only a one-to-one buyer-seller relationship is possible or desired, virtual private marketplaces or B2B collaboration networks preserve pre-negotiated terms and relationships between specific buyers and suppliers so that their members can interact privately in the kinds of detailed, higher-value interactions they have always done off-line and are now pursuing on the Internet. B2B collaboration networks provide an integrated software platform that connects companies in specific industries and mainly facilitate the procurement and sales functions of the collaborative commerce network. Thus, these private marketplaces let their members share privileged information within a defined community of trusted business partners. They are mostly found in situations involving mission-critical, high value inputs and components or where high fixed costs occur.

For these situations, i.e., when only long-term sourcing is available and when only private marketplaces are used, the rich literature on traditional supply contracts applies. However, these models on supply contracts provide little insights on how the possible presence of a second procurement/sales channel of short-term nature affects the situation. The supply chain contract literature also ignores whether there will be a co-existence of long-term and short-term contracts or whether only one of these contract types will prevail. Later in this chapter, we present a quantitative model that addresses exactly this issue.

Note that it is possible to consider second (spot) sourcing under some conditions (though our initial description in Figure 3.2 precluded this possibility), even for nominally customized parts. This is a consequence of the fact that a customized part, such as a forging, normally considered unique, could in fact be spot sourced by spot *capacity* that has been obtained, and is used either with a spare identical die, or with the original die, while the original capacity is being used to produce another part type.

2.2 Catalog Hubs

Catalog hubs create value to all players by automating sourcing and procurement processes and reducing transaction costs when the details of the transaction have previously been agreed upon. They are designed to support systematic purchases when transactions take place with pre-qualified suppliers using contractually agreed upon business rules, e.g., pre-negotiated prices or pricing schedules. Catalog hubs can also aggregate demand or supply in markets with fragmented buyers or sellers and create value when there are a few large players and many small players of the same type (buyer or seller). Note that in contrast to Kaplan and Sawhney (2000) we do not limit catalog hubs to industry specific manufacturing inputs but rather define them, more generally, for the systematic sourcing of goods and services, which are traded according to pre-defined business rules.

With systematic sourcing, and standardized components, it is possible to have alternate pricing schemes, when prices of components may not necessarily be fixed. This corresponds for instance to the situation when the manufacturer and the supplier share risk and reward with respect to demand uncertainty that results in unused capacity or unmet demand. One such scheme is based on linear pricing, where the actual price is lowest when all the committed capacity is acquired, but increases linearly, as a smaller fraction of the committed capacity is actually acquired/used by the manufacturer. We discuss this mode, together with the next one (spot market), in the bulk of the chapter, as the alternate modes of most interest to us from an overall economics point of view of the B2B exchange/marketplace. It is important to recognize again that although catalog hubs by themselves may have similar characteristics to private marketplaces, the goods or services traded require a completely different modelling approach as they are sufficiently standardized, meaning that spot markets can be established for them, too.

2.3 Spot markets

Close cousins of traditional commodity exchanges, spot markets or on-line exchanges allow managers to smooth out peaks and lows in demand and supply by rapidly exchanging the good/service needed or produced. It is important to notice that the type of good/service traded on exchanges can also include flexible production capacity, labor, transportation, and advertising resources.

The nature of exchanges implies that the pricing structure on these exchanges is dynamic and that buyers will thus face price uncertainty as supply and demand may change at any given point in time. This type of B2B e-commerce adds the most value in situations with a high degree of price and demand volatility, e.g., in utilities markets, high fixed cost assets with long acquisition lead-time, such as manufacturing capacity, or perishable items, such as transportation capacity and food.

While there are several types of market-making models, those most pertinent to spot markets and the current chapter are the auction model and the exchange model.

- Auction Model

Auction models are most appropriate in industries where one-of-a-kind, non-standard, or perishable items or services are traded among businesses that have different perceptions of value for the item. Capital equipment, used products, and hard-to-find items fit this model. However, of more direct interest in the context of the current paper are multi-lot auctions and repeated auctions, which can help generate equivalent spot market demand-price curves. The fact that similar (or same) quantities may have different prices in different auctions results in demand and random price curves.

- Exchange Model

Exchange models create value by temporal matching of supply and demand using real-time, bid-ask matching processes and market wide price determination. The exchange model is suited best for near-commodity items with volatile demand. Similar to the situation in financial markets, derivatives such as forward contracts and options, can be devised for exchanges or spot markets, with respect to commodities or flexible manufacturing capacity. This market type is closest to the theme of the current paper, which is the direction that electronic markets are moving in.

3. Characteristics of Markets Participating in B2B E-Commerce

There are several characteristics of different sources (markets) which impact supplier (source) risk management when these sources (markets) are interacting in a B2B exchange, which we will discuss below. In the later parts of the current chapter, for pedagogical reasons, we only describe the results corresponding to sources/markets which possess the least complicated characteristics of those we discuss below, e.g., perfect quality of supply.

1. Liquidity of Markets

Liquidity has two essential components: participant liquidity and transaction liquidity. In our context, the latter is of greater importance, which can be interpreted as transaction volume. A detailed analysis reveals the fact that almost all of the existing marketplaces are currently struggling with low transaction volumes. Only exchanges which are formally associated with large players or led by an industry consortium guaranteeing a minimum transaction volume, have been able to overcome this problem. Limited liquidity may lead to situations where demand (temporarily or in general) outstrips supply meaning that the capacity of the exchange is not sufficient to satisfy all of the demand, an issue that is incorporated in one of the models we present.

2. Quality of Supply

To ensure a minimum quality of supply and delivery performance, market-makers are setting up qualification processes for buyers, sellers, and the goods that trade hands and monitor their ratings, e.g., SupplyWorks teamed up with Open Ratings to build a supplier-performance rating system that customers can access using supply chain marketplaces and exchanges built on SupplyWorks' Max e-procurement solution. Companies make also increasing use of third-party inspectors who perform physical inspection of goods that trade hands to reduce supply quality uncertainty. Despite all these measures, sources/markets have varying degrees of quality, and this characteristic is certainly a key element when managing economic risk in drawing upon one or more sources and meeting demand (sometimes with a minimal quality requirement). Later in this chapter, again for pedagogical reasons, we focus on perfect quality sources, and only briefly allude in passing to the implications of differing levels of quality on the nature of "optimal" polices and the resulting costs and/or profits.

3. Channel Conflicts

Although existing intermediaries and new B2B exchanges can act as complements and are not necessarily substitutes, it is still unclear how spot markets affect existing buyer-supplier relationships. For instance, for electronic parts distribution, Avnet and Arrow Electronics, two dominant incumbents, joined forces with ChipCenter and QuestLink to form eChips in order to compete with the newly created marketplaces E2open (a consortium of IBM, Hitachi, Nortel Networks, Toshiba, Lucent, and Solectron) and eHitex (created among others by Compaq, HP, and Gateway). Thus, as an alternative to the collaborative framework used in this chapter, where we consider a cooperative approach to use a long-term (reservation capacity based) supplier and a spot market, there could be a multiple channel environment, where the channels are in conflict. Dynamic game models can be used to model these environments.

4. Pricing Dynamics

While fixed prices are well suited for small-ticket items with small transaction volume or catalog models with pre-qualified suppliers and predefined business rules, price mechanisms in exchanges, in particular in spot markets, using real-time, bid-ask matching processes will result in highly dynamic prices. Not surprisingly, most procurement managers do not know how to deal with the new volatile price environment and are thus reluctant to fully use B2B exchanges.

Moreover, as the implicit value of any supply contract is driven by the underlying demand distribution for the item in question, the perceived value of a contract differs from company to company as it is driven by (among other factors) the distribution of a company's end demand. Hence the question arises on what a fair and efficient price, e.g., for perishable capacity, should be. Similar to approaches for financial exchanges, one could argue that there should exist a price equilibrium between long-term and short-term contracts. However, the argument becomes much more involved as the value of a contract is potentially different for each of the players. Furthermore, from a supplier's point of view, dynamic pricing bears the risk of alienating big, strategic customers who insist on lower prices than their competitors or other customers of the supplier and who may find out that a component is sold at a lower price via a B2B exchange.

Later in this chapter, we use a simple affine pricing schedule for the reservation capacity cost as a function of the quantity actually bought in relation to the quantity reserved initially. Consequently, this pricing schedule captures the risk sharing agreement between a manufacturer and his long-term supplier for the supplier's underutilized capacity resulting from the mismatch between the reserved and used capacity. The spot market is assumed to possess a fairly general random price that may depend on the amount the manufacturer purchases on the spot market.

5. Fairness

One of the most commonly cited challenges for B2B e-commerce is the question of fairness in B2B exchanges. As stated above, the implicit value of any supply contract depends on the underlying demand distribution, and consequently a contract may have very different values to different companies. Hence it is unclear what a fair price for the item traded should be (in particular, perishable items or goods with high holding and storage costs will cause difficulties). Similarly, some benefits due to exchanges are only the result of the interaction of several players, e.g., it is widely accepted that demand aggregation reduces purchasing costs through quantity discounts. Although it seems intuitive that the savings due to demand aggregation are passed on such that bigger buyers obtain a higher share of the total savings, it is not clear how this issue is going to be resolved. That is, there are no models that are widely used in practice to compute optimal solutions, despite the fact that the problem is a long standing one that is well known in the economics literature.

6. Incentives

A recent report by AMR Research (2000) states that suppliers balk at using Web exchanges, especially since they pay the fees charged when doing business on-line. Meanwhile, suppliers must still manage their own inventory, logistics, and customer service. Overall, most experts agree that there must be value-added propositions for both buyers and suppliers to induce membership in any specific exchange. "If these exchanges are to succeed, there has to be a superior value opportunity for all sides, and if not, then no side will participate," says Chuck Donchess, executive vice president and chief strategy officer for Commerce One. Consequently, some modifications may be required in the future, in adapting the results in this chapter, so that the supplier's economic perspective is captured in implementing risk management from a manufacturer's perspective.

7. Spot market with Lead-time and Derivative Prices

A spot market can have either zero or finite lag (lead) time. Similarly, the demand-price curve can have a fixed price, random price, or the price can be a derivative such as a forward price, real option, etc. In this chapter, we consider spot market prices as fixed or random, assume zero lead times, and defer a discussion of the other variants to another paper.

8. Multiple Parameter Trade-offs and Risk-preference Structures

Having access to more alternatives increases the complexity of decision-making to both vendors and customers. Both will have to trade-off multiple criteria, such as price, quality of supply, or lead-time. For instance, some spot markets may offer the same item at a lower price, but also with a lower quality of supply than others at the same point in time. Furthermore, although firms should be risk-neutral with respect to small investment and procurement decisions, many managers would like to account for their aversion against the risk associated with exchanges. Both the multi-parameter trade-offs and risk preferences are not explicitly considered in this chapter, for ease of exposition.

The extensive list of issues with regard to B2B e-commerce implies that the associated decision-making becomes very complex and that there is an urgent need for decision support tools accounting for uncertainties and risks associated with B2B e-commerce. To this end, in this chapter, we will describe recent novel research which provides answers to some of the procurement issues discussed above. We begin our presentation with a short review of traditional supply contract literature which applies to private marketplaces. We then continue with more recent research that mainly focuses on exchanges, and more specifically, on risk management through the interaction of (spot market) exchanges and traditional supply contracts. This work is of great significance in helping manufacturers manage their economic risk in choosing between different types of suppliers such as:

- a possibly lower priced long-term (reservation) capacity, with a risk of having to pay for unused capacity if demand is low (that can be shared with the supplier), versus
- a (potentially) more expensive spot source, which ensures excess demand is met in the instance that demand is significantly higher than anticipated and provided for with the long-term supplier alone.

In the next section, we begin with an overview of some recent research which addresses the quantitative risk management issues we have repeatedly alluded to above and then describe the results in some detail, and conclude with future directions.

4. Quantitative Models for B2B E-Commerce

We now describe the concrete results available to address the issue of economic risk management in supplier management in a B2B exchange, including our most recent research (Araman et al. (2000a), Araman et al. (2000b)), to which we will refer. Each of these sections concerns research that develops solutions for each of the markets in the boxes of Figure 3.2. We move from results concerning private markets, to those concerning spot markets, which are of particular interest to us.

4.1 Traditional Supply Contract Literature (Focus: Private Marketplaces)

The literature on production-inventory systems, and particularly, supply chain contracts, both game and non-game based models, has exploded over the past few years. The results are very pertinent to risk management in private markets and provide useful insights for one-to-one type supplier-customer relationships or other forms of long-term interactions. Two recent review papers (and Bassok and Anupindi (1997b), as well as Tsay et al. (1998)) summarize many of the results. The research on supply contracts can be fundamentally separated into supply contracts with no commitments and supply contracts with commitments (regarding order quantity or delivery performance). Bassok and Anupindi (1997a) analyze a single product contract where the supplier offers discounts for a total minimum quantity commitment to a buyer facing stochastic demands. Anupindi and Akella (1993) consider another form of commitment, in restricting the periodic order quantity of the buyer. They study a class of contracts that require the buyer to commit, at the beginning of the planning horizon, to purchase a certain minimum quantity in every period. Orders can be adjusted upwards for payment of a price premium. Furthermore, the contract reflects the delivery responsiveness of the supplier to the order quantity adjustments. Observe that these papers address the issue of trading off upside and/or downside risk through operational policies, but do not directly incorporate reservation capacity related pricing.

Similarly, a number of recent papers study the effect of dual sourcing on a production system. Although dual sourcing can be interpreted as an alternative framework to the one presented in this chapter, it does not fully capture the effect of capacity reservation and the respective literature focuses more on inventory related issues. For instance, Fong, Gempesaw, and Ord (2000) as well as Rudi (1999) analyze the effect of dual sourcing on the optimal inventory. Fong et al. study an inventory system with a choice between two supply options having different lead-times. They allow for normally distributed demand and Erlang distributed supplier lead-times. In contrast to Fong et al., Rudi does not assume that orders for both supply options are placed concurrently and considers sequential decision-making. He uses a two-stage stochastic linear programming formulation to analyze the split between make-to-stock and assemble-to-order. Both papers assume that all cost parameters are deterministic.

4.2 Simulation based Supply Contracts, Applied to a Private Market, and a Spot Market

As described above, the traditional supply chain literature does not account for the recent development in B2B e-commerce, especially with regard to Internet-based exchanges. Thus, there is an urgent need for new models that address B2B exchanges and their impact on current procurement practices. One can argue that the complexity of the problems, the number of parameters involved and their characteristics make simulation techniques a useful approach for such problems.

The recent paper by Cohen and Agrawal (1999) solves a stochastic dynamic programming formulation using simulation to analyze the trade-off between long- and short-term contracts. Their model, however, only allows for the usage of either contract at a time, i.e., a mixed strategy is not considered. The model proposed in this paper can be used for the comparison of long-term contracts and a spot market since a spot market can be approximated by a short-term contract with a very short duration. We discuss this in more detail below. The elements that dominate are the uncertain prices of the spot market versus fixed investment costs and learning cost reductions by the long-term supplier, along with the usual inventory and backlog costs.

As mentioned above, the paper by Cohen et al. proposes a simulation-based model comparing short and long-term contracts. They consider a planning horizon of several years divided into T_M review periods. Each review period is, in general, of one year in length or divided into T_T tactical review periods (a tactical review period is 1 week long and T_T is thus equal to 52). If the long-term contract is selected at any point during the horizon, it lasts for the remainder of the horizon. The short term contract however lasts, if selected, for the duration of a review period. During each review period the system evolves according to the selected contract. We define y_t as a 0-1 decision variable (0 for short term and 1 for long-term).

The model and the results are valid for a shorter review period, with T_T smaller than 52. The only assumption needed is that the system reaches stationarity during one review period. Hence by shrinking the length of the short term contract as defined in Cohen et al. the short term contract could be replaced by a spot market as we defined it earlier in this work, both having almost the same characteristics as far as the model is concerned. Thus, we are interpreting the conclusions of the paper in our framework, so that the results and insights can be used in a B2B exchange framework. The supply managers will still face the same dilemma of choosing between either a short term or a long-term contract. The first option has the advantage of offering the flexibility to switch to different suppliers and is based on a speculative (or random, in the sense of not known a priori) market price, whereas the long-term relationship requires a fixed initial investment and is based on a price written into the contract. In addition, learning effects due to the long-term relationship are described by an annual percentage reduction, δ, of the total cost incurred.

This paper considers a risk averse supplier whose problem is to select the optimal contract for each strategic period by minimizing the total disutility of costs over the entire horizon. To account for the buyer's risk aversion, the authors suggest that the buyer follows a "mean-variance" type of function to evaluate the total cost *C,* incurred in one review period:

$$U(C) = E[C] + \lambda \mathrm{Var}(C)$$

where λ ($\lambda \geq 0$) reflects the risk aversion of the buyer. (The more risk averse the buyer, the larger the value of λ.) Hence, the problem is formulated as a stochastic dynamic program, where at each review period the manager chooses one of the two types of contracts for that period, based mainly on the previous price of the spot market. He will then need to decide on the tactical inventory policies. That is, he has to account for the cost of goods (P_n, total purchase price paid during period n), the fixed investment when the long-term contract is selected, *K, as* well as the inventory and shortage costs. The optimization is performed over the entire horizon. Although the main focus of the paper is on the non-stationary, finite horizon setting, a first intuitive result shows that under the stationary, infinite horizon assumption (i.e., the spot market price is time stationary) a "now or never" kind of policy is adopted; that is, the long-term contract is selected either at the very beginning or never. The reason is that the formulation of the problem is the same every period and is independent of the other periods. As a consequence of the non-stationarity, the formulation of the problem becomes similar to that of American call options for which the solution can only be determined numerically.

In what follows we will briefly describe the different parameters that are included in the model, present a summary of the results, and conclude with some managerial insights. For a more detailed description of the model the reader is referred to Cohen and Agrawal (1999).

As we stated earlier, the spot price plays a major role in this model, and its non-stationarity is crucial. The distribution of the spot market price is assumed to follow a multiplicative Binomial process, and to only depend on the realization of the previous period. We let $\tilde{p}_n$ denote the random market price and $\hat{p}_n$ the realization of the price at time n. Then

$$\widehat{p}_{n+1} = \begin{cases} p^u_{n+1} = \Delta^u \cdot \widehat{p}_n & \text{with probability } \pi^u = 0.5 \\ p^d_{n+1} = \Delta^d \cdot \widehat{p}_n & \text{with probability } \pi^d = 0.5 \end{cases}$$

where Δ^u and Δ^d are two constants that determine the upward and downward trend of the price. π^u and π^d are usually assumed to equal 0.5. This results clearly in a non-stationary environment where the prices in successive periods are correlated. On the other hand, the price specified by the long-term contract at a given point in time is locked in for the remainder of the horizon, and is assumed to equal the spot market average price during this period. We note that the selection of the contract occurs before the spot price is revealed in this particular period. The total expected purchase cost during one review period is

$$P_n(y_n) = \mu T_T p_n(y_n)$$

where μ is the average demand and

$$p_n(y_n) = \begin{cases} \tilde{p}_n & \text{if } y_n = 0 \\ \pi_u p^u_n + \pi_d p^d_n & \text{if } y_n = 1 \end{cases}$$

In addition, the supply manager will also decide on an order (up to) policy, by trading off holding and penalty costs. This results in a quasi news-vendor model that incorporates the different lead times based on the contract's type and the length of the horizon. Let $C_n(y_n)$ be the aggregate inventory and shortage costs. The dynamic programming formulation is then given by

$$U(\nu(y_{n-1}, \hat{p}_{n-1}) = \min_{y_n} \{K(y_n - y_{n-1}) + P_n(y_n) + C_n(y_n) + \zeta U(\nu(y_n, \tilde{p}_n))\}$$

where ζ is the discount factor. The simulation-based conclusions are very insightful and interesting. As stated above, the long-term contracts are characterized by the rate of cost improvement, δ, and the first conclusion is that long-term contracts are worth considering if δ is higher than a threshold value so as to compensate for the high initial investment, *K,* incurred. This partially explains why managers are reluctant to establish a long-term relation with a supplier. A "wait and see" policy is usually adopted in a non-stationary environment where usage of the short term contract (or use of the spot market in our case) is more suitable at the beginning; then, based on the dynamics of the spot price, as described by the drift and volatility of the price process, the decision is made to either wait longer or to lock in the price in a long-term contract. This strategy clearly depends on the initial price distribution known to the managers, as well as their risk aversion. If the probability that the price will go up is high and the decision-maker is highly risk averse, a long-term contract is selected in some cases from the very beginning, even with high initial fixed investment, *K.* The main contribution of this work is to identify, in fairly general settings, the factors that contribute to a contract selection. These factors are the fixed investments, the length of the planning horizon, the improvement rate for long-term contracts, the risk aversion of the decision maker and finally the price uncertainty of the spot markets.

4.3 Supplier Procurement Risk Management

The previous approach, although very comprehensive and broad, did not allow for a simultaneous use of both modes of supply: the long-term contract and the spot market. Introducing this option into the model is not just a means to complete the previous results but a way to study how the presence of the spot market affect the traditional supply channels, i.e., long-term contracts, whether one dominates the other, or whether they could possibly coexist together. As we stressed earlier, introducing the various characteristics of those contracts makes the problem analytically intractable. For instance, a major difficulty introduced in the Cohen and Agrawal model was the non-stationarity of the spot market price. Although this factor could be of primary concern, relaxing it should enable us to get the right insights concerning how the two methods of procurement evolve together. Once we make the stationarity assumption, which corresponds to an equilibrium stage, the setting is equivalent to a one-period problem, as shown above, because the state space is then independent of time. What is optimal in a particular period will remain optimal in a subsequent period, i.e., a myopic policy is optimal. In the following paragraph, we describe two models addressing the question of whether a spot market and long-term contracts will coexist or not, with many similarities but several significantly complementary features. The first one considers a general setting with non-linear pricing for both a long-term supply mode and a spot market while still allowing us to draw analytical conclusions on the optimal procurement strategy. The second model assumes specific distributions of the spot price and the demand, but it identifies in some cases closed form solutions of the optimal procurement strategy, and studies the sensitivity of the model, in general, to the different factors involved, including the risk preference of the decision maker. In any case, both models show the optimality of a mixed strategy in such an environment, i.e., both modes of supply will coexist.

4.3.1 A General Approach for a Risk-Neutral Decision-Maker. The main objective of this model by Araman et al. (2000b) is to determine the optimal strategy that the buyer should adopt (in the following we will interchangeably use buyer and manufacturer) when two main modes of supply are available: a procurement channel through the spot market and a long-term channel defined by a reservation capacity level. In line with the widely known fact that corporations should be risk-neutral for decisions involving small investments relative to their overall wealth (which is true for most procurement decisions), we model the manufacturer as a risk-neutral decision-maker. Furthermore, to reflect the recent trend towards customized products and just-in-time manufacturing, as manifested in the automotive industry, we model the manufacturer's production environment as make-to-order (MTO), which has the additional benefit of making our model easier to follow. The manufacturer has to satisfy an aggregated (end-market) demand, d_m, via the two procurement channels. The long-term contract, on one hand, specifies a unit price schedule and a capacity level that is reserved for the manufacturer. The spot market, on the other hand, is character-

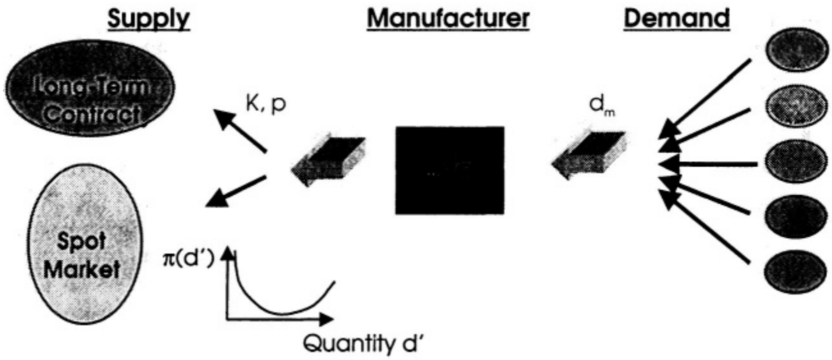


Figure 3.3. Manufacturer (a) and Channel Profit (b) - Case 1

ized by a price-quantity curve $\pi(d)$. Figure 3.3 describes the conceptual framework, detailed in the next four subsections.

In order for the manufacturer to achieve her serviceability objective and hedge against the variability of demand and supply (which may also be impacted by the supplier's capacity limitations), the manufacturer offers her supplier a long-term contract defined as follows: In period 1 the manufacturer will determine a capacity reservation level *K,* based on a pricing scheme $P_K(d)$. Subsequently, in period 2, the manufacturer has to meet a random demand d_m. When the demand is lower than *K,* $(d_m < K)$, she will order d_m from the supplier and pay a unit price of $P_K(d_m)$. Similarly when the demand is higher than *K,* $(d_m \geq K)$, the manufacturer will order the maximum reserved capacity, i.e., *K* units, from the supplier at a unit price of $P_K(K)$. In this case, however, in order for the manufacturer to meet her total demand, she will need to procure the remaining units from the spot market. We assume that the total demand d_m has a continuous density function f_{d_m} and admits finite first and second moments.

The Supplier Pricing Scheme. To allow for risk-sharing between the long-term supplier and the manufacturer, the long-term supplier will charge a unit price $P_K(d)$ for an order of d units. Clearly, this price depends on the capacity that the manufacturer reserved in the original contract, *K,* and on her actual order, d. If the reserved capacity is fully used, the supplier will charge a unit price of $P_K(K)$. If only an amount d with $d < K$ is used, he will charge $P_K(d) > P_K(K)$ to account for any losses due to the under-usage of reserved capacity. Hence $\xi_K(d) = P_K(d) - P_K(K)$ is the capacity reservation penalty. The assumptions on the pricing scheme are the following:

- For all K the penalty ξ is linear in d. This is similar to the approaches by Li and Kouvelis (1999) as well as by Barnes-Schuster et al. (1998).
- The cost of goods $P_K(d) \cdot d$ is always increasing with d, i.e., *it costs more to buy more.* As demand is random, the manufacturer and supplier commit to a risk-sharing agreement in the following sense: the penalty in total supply cost for "under-ordering" with respect to K becomes smaller as the order quantity approaches K.
- $P_K(K) = p$ is independent of K. This assumption is for pedagogic reasons only. However, it can be interpreted as the result of the negotiation between manufacturer and supplier. The first wants to have $P_K(K)$ non-increasing in K, whereas the latter prefers to see $P_K(K)$ being non-decreasing in K, which may lead to $P_K(K) = p$, independent of K. (See Araman et al. (2000b) for a more detailed justification and a model without this assumption).

  Based on the previous assumptions, a linear function that fulfills all of these conditions has to be of the form:

$$P_K(d) = p + \xi_K(d) = 2p - \frac{p}{K}d \text{ for all } d \leq K \qquad (3.1)$$
$$P_0(d) = 0$$

  Thus the penalty ξ_K is given by: $\xi_K(d) = p(1 - \frac{d}{K})$. See Figure 3.4 below for an illustration of the long-term pricing schedule $P_K(d)$.

Remark 3.3 *We should note that we could relax the last assumption related to $P_K(K)$ from being constant (of value p) to just being non-increasing with K. That is, the results to follow in this section will remain valid for any pricing scheme that fulfills the first two assumptions above (i.e., regarding the linearity of the unit cost and the increasing property of the total cost). The choice of (3.1) is only for clarity of exposition.*

The Spot Market. Companies expect overhead to decline and prices for materials and components to drop dramatically in the short term as suppliers are forced to compete head-to-head on-line. However, we believe that once equilibrium is re-established, prices for long-term contracts will tend to be lower than spot market prices as the additional flexibility on the spot market will be rewarded with a price premium, at least in expectation. Furthermore, as the spot market for most industries is still maturing and there is little data or analysis available

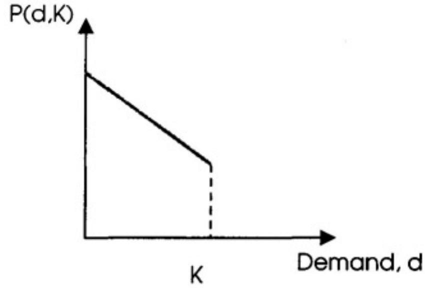


Figure 3.4. Long-Term Pricing Schedule

that describes the behavior of these markets precisely and since industry experts predict that different market mechanisms will co-exist on each exchange, we have exercised great caution in making assumptions about the spot market and define it as a random price-quantity function. Thus the unit price $\pi(d)$ is a random variable that depends on the number of units d ordered, such that for a fixed d, π is defined by its density function $f_{\pi(d)}(.)$. π is in addition dependent on and most likely positively correlated with the total demand d_m received by the buyer. Since we consider the optimal mix with respect to the expected value of the procurement costs, it is sufficient to consider that the spot market price is given by $\hat{\pi}(d, d_m) = E[\pi(d)|d_m]$ where for every order d and a total end market demand d_m, $\hat{\pi}(d, d_m)$ is the expected value of the price. In addition, as we did for the long-term supply pricing scheme, we assume that the function $\Pi(d) = \hat{\pi}(d, d_m) \cdot d$ is increasing in d, i.e., the total cost of goods increases with the quantity of goods bought from the spot market. We let $\Pi(0) = 0$. On the other hand, as we stressed in the first sections of this chapter, one of the main issues related to spot markets is their liquidity, as manifested by the current low transaction volume of most spot markets. That means that companies may not be able to fill all of their demands using only the spot market. We therefore consider in the general case a spot market with a capacity constraint, i.e., C is the maximum amount available at the time when the buyer places her order. We assume that C is random, possibly dependent on the total demand d_m, and bounded by a constant value C_∞.

For clarity of exposition we make the assumption that π is independent of d_m and that C is constant (possibly infinite). Again the main results hold true in the general case (see Araman et al. (2000a), and Araman et al. (2000b) for details). Therefore, unless stated otherwise, we consider $\pi(d) = \hat{\pi}(d) = E[\pi(d)]$. We finally assume that:

$$\overline{\Pi} = \int_{-\infty}^{C} \Pi(u) f_{d_m}(u) du < \infty$$
$$\overline{\Pi}' = \int_{-\infty}^{C} \Pi'(u) f_{d_m}(u) du < \infty$$

Problem Formulation. Our first approach is to consider a risk-neutral buyer interested in minimizing her expected total procurement cost $G(K)$. As we will see below, the model is defined such that higher moments of the demand distribution will come into play. Thus, despite the assumption that the decision-maker is risk-neutral, which makes the problem tractable, the model still captures the volatilities of both the demand and the spot market. If we denote $\Psi(K)$, the total random cost associated with a reserved capacity level K, the optimization problem of the manufacturer can be written as follows: (We will use the following notation: $E[X;A]$ that should be read $E[X \,.\, I(A)]$, where I is the indicator function. It means the expected value of the random variable X, under the event A, which is not to be confused with a conditional expectation.)

$$\min_K G(K) = \min_K E[\Psi(K)] = \min_K \{E[\Psi; d_m < K] + E[\Psi; d_m \geq K]\} \tag{3.2}$$

When the demand of the manufacturer is less than K, she will only use the long-term contract, and therefore:

$$E[\Psi; d_m < K] = \int_{-\infty}^{K} u P_K(u) f_{dm}(u) du \tag{3.3}$$

However when the demand is higher than K, exactly K units will be supplied by the long-term contract and an additional $d_m - K$ units will have to be purchased on the spot market. Again, since most spot markets are currently struggling with low transaction volumes, the manufacturer may not be able to purchase as many units on the spot market as needed and the amount available to the manufacturer is given by $\max(C, d-K)$, where C is the spot market capacity. We assume that the manufacturer incurs a fixed penalty cost per unit, b, on those units that she could not buy on the spot market.

$$
\begin{aligned}
E[\Psi; d_m \geq K] &= \int_K^{+\infty} Kpf_{dm}(u)du + \int_K^{C+K} (u-K)\pi(u-K)f_{dm}(u)du \\
&\quad + bE[d_m - (K+C)]^+ \\
&= Kp(1 - F_{dm}(K)) + \int_K^{C+K} \Pi(u-K)f_{dm}(u)du \\
&\quad + bE[d_m - (K+C)]^+
\end{aligned}
$$

We will divide the total expected cost into two parts, slightly differently than just computed. The first term, H_{LT} represents the expected cost that is spent via the long-term contract, whereas the second, H_{SM}, is the expected cost of purchases on the spot market, including the penalty cost for a possible undersupply. We rearrange the terms of the previous equations to obtain:

$$
\begin{aligned}
H_{LT}(K) &= \int_{-\infty}^{K} uP_K(u)f_{d_m}(u)du + Kp(1 - F_{d_m}(K)) \\
&= \int_{-\infty}^{K} p(2 - \frac{u}{K})uf_{dm}(u)du + Kp(1 - F_{dm}(K)) \\
H_{SM}(K) &= \int_K^{C+K} \Pi(u-K)f_{dm}(u)du + bE[d_m - (K+C)]^+
\end{aligned}
$$

and the total expected cost is again the sum of the two terms:

$$
G(K) = H_{LT}(K) + H_{SM}(K)
$$

Proposition 3.4 *$H_{LT}(K)$ is a positive, increasing, and concave function of K such that $H_{LT}(0) = 0$ and H converges asymptotically to a finite constant $H_{LT}(\infty)$ as $K \to \infty$.*

Proposition 3.5 *$H_{SM}(K)$ is a positive decreasing function of K such that $H_{SM}(0) > 0$ and H_{SM} converges asymptotically to* 0 as $K \to \infty$.

These two last propositions are very intuitive. As K increases the (expected) amount paid to the long-term supplier increases and, simultaneously, the amount spent on the spot market decreases in expectation. See Figure 3.5 for an illustration of H_{LT} and H_{SM}. Notice that in the graph below, $H_{SM}(0)$ is greater than $H_{LT}(\infty)$ or in other words, it is more expensive in expectation to only purchase on the spot market than to go exclusively with the long-term supplier.

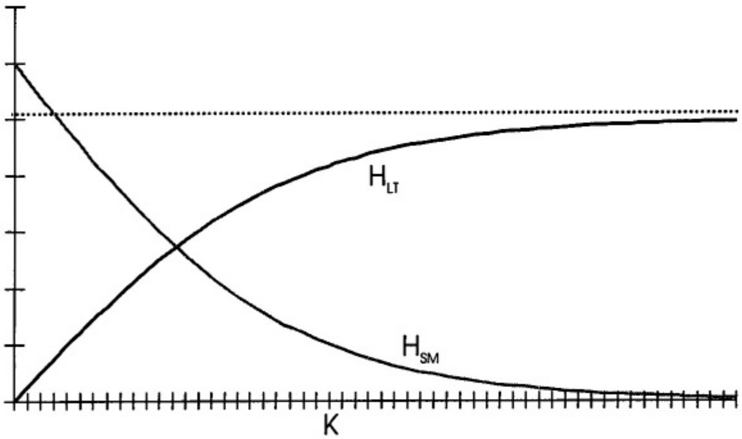


Figure 3.5. Illustration of H_{LT} and H_{SM}

Main Result. We now describe a set of results that indicate that the spot market is valuable under fairly general conditions, and also characterize the values of the (optimal) contracted capacity *K*.

Theorem 3.6 *Let K* be the optimal value of K that minimizes (3.2) and let* $\phi(u) = \ln(f_{d_m}(u))$.

- *If the following assumptions hold:*

$$\exists\ u_0, \forall u \geq u_0 \ and \quad \forall t,\ \phi(t+u) \leq \phi(t) + \phi(u) \tag{3.4}$$

$$\overline{\Pi}' > p \ or \ \overline{\Pi} > 2pE[d_m] \tag{3.5}$$

  then the cost function G admits a minimum at K such that* $0 < K^* < \infty$.

- *If the total demand* d_m *has finite support* $[a, b]$ *with* $0 < a < b < \infty$ *and*

$$\pi(u) > p \ for\ all\ u \in [a, b] \tag{3.6}$$

  then the cost function G admits a minimum at K such that* $a < K^* < b$.

It is important to recognize again that $K^* = b$ ($b = \infty$ in the first case and b finite in the second) corresponds to the case when the manufacturer uses exclusively the long-term supplier and that *K** = 0 is equivalent to the manufacturer only purchasing from the spot market. A value of *K** equal to an interior point of the support of d_m means that a mixed strategy is optimal. Before giving an intuitive interpretation of the previous main result, we will note two things. First many density functions have the characteristics given by (3.4) such as an exponential distribution or a normal distribution; more generally, any random variable with an exponential tail distribution would work. Secondly, (3.5) is basically saying that the spot market is somehow more expensive than the long-term contract. One of these two conditions is enough for the result to hold. The first one is equivalent to saying that the expected marginal cost of procurement from the spot market, $\overline{\Pi}'$, is greater than the minimum long-term procurement cost, p. The second condition means that it is cheaper for one to select the long-term contract as the only means of supply than using the spot market exclusively. In the context of this paper this seems a very reasonable assumption: the exchanges, as we noted earlier, create value to the buyers (e.g., the flexibility of switching to other suppliers or offering great quantity flexibility) and that comes at a certain cost.

Remark 3.7 *By concluding that K* never takes on one of the extreme points of the domain, Theorem 3.6 demonstrates that for a wide range of demand distributions a mixed strategy is always optimal, i.e., the buyer will always choose to meet part of her demand with a long-term supplier (or a catalog hub as defined earlier in the B2B E-commerce framework) and the remaining from spot markets.*

Remark 3.8 *The intuition behind the result of Theorem 3.6 (as illustrated in Figure 3.6) is that the realizations of the higher demand values occur only with a small probability, and thus the manufacturer is better off paying a high price on the spot market for these "rare" ("low probability") events than to almost always pay a penalty for under-using the supplier's capacity. Nevertheless, it is surprising that a mixed strategy is optimal almost independently of the spot market price (e.g., even for a very high one in comparison to the long-term contract pricing schedule) and of the spot market capacity constraint.*

Approximating Bounds and Numerical Solutions. Figure 3.7 illustrates that the optimal mix between long term contract and spot market can lead to a significant reduction in procurement costs in comparison to an exclusive sourcing from the long term supplier (as is the current practice in many industries). In this example, we modeled demand, d_m, as a normal variable and the spot price as $\pi(u-K) = \frac{1}{\sqrt{u-K}} + \pi$. Other functional forms of $\pi(u-K)$ are conceivable, but yield mostly similar results. Notice also that the optimal value of capacity

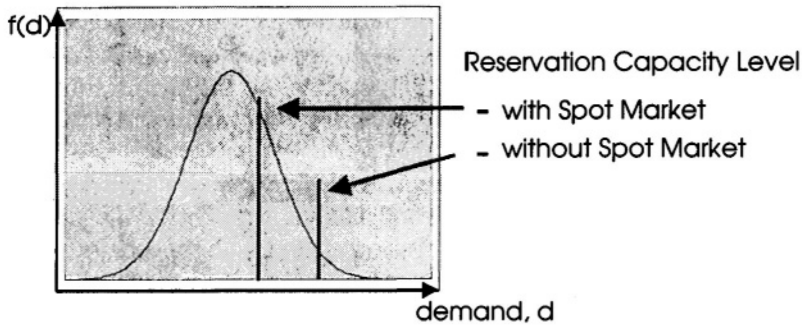


Figure 3.6. Intuition behind Theorem 3.6

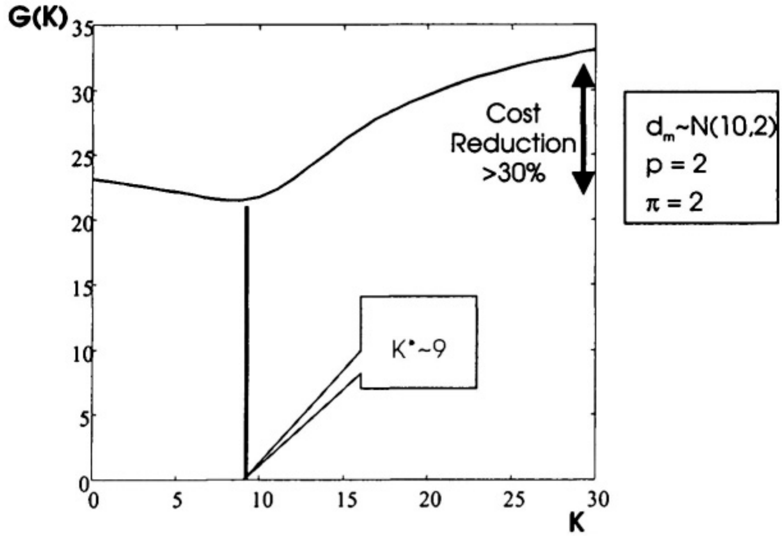


Figure 3.7. Illustration of $G(K)$ as a Function of K for Normal Demand, d_m

reservation with the long-term supplier, K^*, is less than the mean value of demand ($K^* = 9$ vs. $\mu_d = 10$). In general, however, we noticed that K^* is close to the mean value of demand, although most often greater than μ_d, depending on the relative prices of long term contract and spot market.

The general setting we adopted makes it almost impossible, except numerically, to obtain the optimal capacity reservation level, K^*. However, it is important at this stage to get a better feeling of the range of values where K^* could lie. For instance, let us assume that K^*, even though finite, is taking on a high value, far from the mean of the demand. Hence, only a very small portion of each manufacturer's demand will be purchased via the spot market. Therefore, under such a scenario, it is legitimate to wonder if it is really worthwhile establishing these spot markets. Fortunately, by solving the problem numerically, we can see that for reasonable spot market prices, K^* takes on values close to the mean and, in some cases, when the spot market is not too expensive, even smaller than the mean of the demand. In addition, although we were not able to get a closed form solution, we are able to compute, in some fairly general settings, upper and lower bounds for K^* that help determine numerically an approximation of the optimal reserved capacity.

For the remainder of this paragraph we follow the derivation of the bounds for K^* in Araman et al. (2000b) and assume that the density function of the demand is unimodal with m denoting the (unique) mode, i.e.,$\forall u \geq 0$: $f_{dm}(m) \geq f_{dm}(u)$. We consider the spot price to be random and dependent on the total demand d_m received by the manufacturer. Furthermore, we assume that a single spot market unit price applies, regardless of the number of units ordered. This is a reasonable assumption for many commodity-like items with small economies of scale effects. Hence, $\pi(d)$ should read $E[\pi|d]$. Let $\pi_0 = \min_{d \geq 0} \pi(d)$. As we said earlier, the spot market price is assumed to be higher than the minimum price p offered by the supplier, and thus $\pi_0 > p$. We define,

$$\underline{K} = \overline{F}_{d_m}^{-1}\left(\frac{p}{\pi_0}\right) \tag{3.7}$$

Theorem 3.9 *Let $\underline{K}$ be given by (3.7) and $\overline{K}$ the biggest solution of the following equation:*

$$f_{dm}(\overline{K})\left[\left\{3\pi(\overline{K}) - 2p\right\}\overline{K}^3 + 2pm^3\right] = \frac{7}{4}pm^3 f_{dm}\left(\frac{m}{2}\right)$$

Then under the conditions of Theorem 3.6 and if the density function of the demand is unimodal:

$$\underline{K} < K^* < \overline{K}$$

Remark 3.10 *The main benefit of Theorem 3.9 is to reduce the computational complexity necessary to determine the optimal level of capacity reservation, K^*. In addition, as illustrated in Figure 3.8, the expected cost associated with both $\underline{K}$ and $\overline{K}$, $G(\underline{K})$ and $G(\overline{K})$, respectively, are close to the optimal value, $G(K^*)$ for small values of standard deviation in demand, σ_d. For larger values of σ_d, $G(\overline{K})$ can be a good approximation of $G(K^*)$, with, in this case, associated error of less than 10%.*

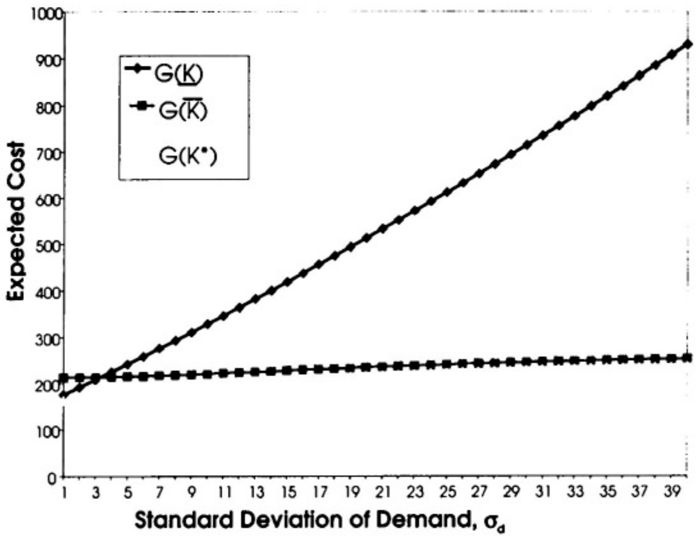


Figure 3.8. Illustration of $G(K^{*})$, $G\ \overline{K}$, and $G(\underline{K})$ as a Function of σ_d for $d_m \sim N(80, \sigma_d)$, $p = 2$, $\pi(d) = 2p\ \ 1 + \sqrt{d}$

4.3.2 Assessment of the Spot Market Advantages for a Risk Averse Buyer. The second model, which focuses on the coexistence of both modes of supply, is due to Seifert et al. (2000). The authors define a one-period problem, where supplies can be procured either through a forward contract or via a spot market. As opposed to the earlier model, the authors in this paper consider constant unit prices (similar to some extent to a newsvendor model) and assume the random variables to be normally distributed. Moreover, they introduce a factor to account for any potential risk aversion of the decision-maker in a similar manner to Cohen et al. (1999) and present a complete sensitivity analysis of the ordering strategy with respect to the different factors involved. Furthermore, they propose that the procurement costs for both the long-term contract and the spot market are linearly increasing in the quantities purchased and they also study different modes of the system where the spot market is used by the manufacturer for buying and selling (BS model) as opposed to buying (BO model) or selling (SO model) only or not using it at all (from the buyers perspective). The buyer in this context is interested in maximizing the profit Π. The long-term contract (defined here as a forward contract) is based on a fixed unit price c and a reservation quantity q with q being the decision variable. As the random demand ξ is received, the buyer will try to meet ξ by using first the q units available from the long-term supply and if more units are needed purchasing them from the spot market at a random price s. On average the spot market is considered more expensive than the forward option, with, however, a negligible lead-time compared to the long-term contract. When the demand ξ is less than q then the difference $q - \xi$ is salvaged at a unit value ν. A fixed unit revenue r is generated from each unit of satisfied demand. When the spot market is used for buying and selling then $\nu = s$, otherwise, the salvage unit value, ν, is taken to be constant, non-random, and smaller than the long-term unit price, c. The analysis is performed assuming positive correlation between the demand and the spot market price, s, and more specifically (ξ, s) is considered as a bivariate normal distribution with correlation $\rho \geq 0$: $(\xi, s) \sim BN(\mu_d, \mu_s, \sigma_d^2, \sigma_s^2, \rho)$. In the BS case, which reflects the fact that the traditional roles of buyers and sellers become unclear in a spot market, the authors obtain a closed form solution for the optimal contract quantity:

$$q^*_{BS} = \mu_d + \frac{\mu_s - c}{2k\sigma_s^2} - \rho\frac{\sigma_d}{\sigma_s}(r - \mu_s)$$

when solving

$$\max_{q \geq 0} E\Pi_{BS}(q) - kVar\Pi_{BS}(q)$$

When studying the sensitivity of these results, the authors find out that as the buyer in the BS model is more risk averse (k positive and increasing) the optimal amount q^*_{BS}, defined in the contract and available to the buyer at the price c, will decrease. This behavior is due to the fact that the spot market is used to salvage the units in excess. Hence, decreasing q will decrease the amount that will be sold at the speculative price s. We note that as $k \to \infty$, q^*_{BS} gets closer to the mean demand and the profit approaches $(r-c)\mu_d$, which is the amount that is achieved without the presence of a spot market. Another important factor is the volatility σ_s of the spot market. Increases in σ_s have a similar impact on the ordering quantity to an increase in k. When the volatility of the spot market is small, the spot price approaches its mean, μ_s, which is higher than c, the unit cost. The decision maker will tend to order a higher value q_{BS}. On the other hand, as σ_s becomes higher, the risk averse buyer would want to use as little as possible the spot market. We could study similarly a Buy Only model (BO), that is characterized by a constant salvage value as opposed to Selling only (SO) where the speculative part of the market enters into account in the salvage value. By comparing all of these models we observe that the buying models (BS and BO) will generate the highest profit in expected value by taking advantage of the volatility of the market, and thus the profits will also have the highest variances.

On the other hand, these two models will clearly increase the fill rate of the buyer, and the existence of the spot markets will create a new source of supply that will help the buyer to better meet the demand. All of these points stress the value that is created by introducing spot markets. The co-existence of the spot market with the traditional forward contracts is another conclusion that was implicit throughout this analysis. This becomes even clearer when studying the numerical examples that show that the ordering quantities via the forward contract are in most cases smaller than or close to the mean demand.

The Supplier's Point of View. The various models described above consider mainly the advantages of spot markets to the buyers. As important as is this perspective, it is also crucial to take the suppliers' point of view and study how these markets could affect their profitability. A first analysis would suggest that the supplier will have access to more customers on the spot markets and therefore one would expect a better management of their capacity due to the diversification of the demand. However, the creation of spot markets, will lead to a highly competitive environment possibly resulting in lower unit prices. Therefore the final outcome is not evident. A preliminary study based on a simpler version of the model of Araman et al. (2000b) shows the existence of a Nash equilibrium between a supplier and a buyer. While the buyer chooses the optimal procurement mix between the spot market and a long-term contract, the supplier has the option of entering a long-term contract with a buyer and/or selling the remaining of her capacity on a spot market. However, more results in this promising direction are needed before we are able to conclude positively on the general economics and value of B2B exchange markets, and spot markets in particular.

5. Outlook and Conclusions

This paper presents a survey of the literature pertaining to analytic approaches for B2B e-commerce. The literature is quite recent, and we note that the research has not yet evolved in a coherent manner nor can it provide answers to all of the issues involved with B2B e-commerce. However, as buyers and suppliers engage in Internet-based marketplaces across industries, the ability to assess the implications of such spot markets and to understand the risks as well as the benefits associated with them becomes crucial. The contribution of this chapter lies in taking an idealistic view of such spot markets, where we show that both spot markets and long-term contracts coexist under fairly general conditions. Knowing that such spot markets can significantly reduce supplier costs, even if the unit price on the spot market is much higher than via the long-term contract, procurement managers should increase their companies' efforts towards Internet-based procurement and prepare its supply chain for such a change. Capitalizing on the various models described above, these managers can effectively manage the risk of such a move and determine the optimal procurement channel mix.

It is important to recognize that the models described here focused primarily on standardized goods and services. However, if customized products are to be exchanged, which are specifically made and designed for one customer, the spot market could be designed to trade production capacity instead of finished components, as long as the same machines can produce several customized products for potentially different companies (possibly requiring a set-up time). The semiconductor industry represents an example where it is common practice to negotiate supplier contracts based on wafer starts per week, i.e., in production capacity. Quality and delivery performance become more important in such a setting, since the manufacturer is highly dependent on the supplier's ability to deliver high quality products on time. However, as more and more Internet-based marketplaces offer quality verification on both the potential buyers and sellers, the main decision-making criteria will be the price of the good offered.

We conclude this chapter by noting a number of specific areas for future research, some of which we already mentioned in our list of B2B market characteristics. Although we made simplifying assumptions for the sake of exposition in this chapter, the papers by Araman et al. (2000a) and Aramant et al. (2000b) address some of the issues listed in the section above, e.g., quality of supply. Furthermore, a forthcoming paper by Kleinknecht et al. (2001) considers portfolios of supply contracts and shows that a manufacturer prefers to access multiple long term suppliers (supply contracts) in addition to the spot market under very realistic conditions. However, despite these recent efforts, the following topics in the discussion of B2B e-commerce have not been (fully) captured by the research literature, and we believe these are very significant research areas in the near future.

- Inventory Levels

Although it is widely expected that B2B e-commerce technologies, and spot markets in particular, will lead to a better utilization of inventories, a model quantifying the extent of improvements is still missing. Existing papers on the benefits of information sharing and the bullwhip effect will have to be re-defined to account for the additional procurement and sales channel through internet-based spot markets.

- Options and Futures

Similar to financial markets or to exchanges in the utility industry, derivatives for the procurement of components and services would be a valuable tool to hedge against uncertainty for both buyer and seller. One of the key differences with respect to financial instruments is that the (perceived) value of contracts and B2B spot markets depends on the demand distribution of the buyer, as well as how mission-critical the component is to the buyer - information which may only be known by the buyer himself (if at all). This raises the issue of what the correct/fair price for a component or unit of production capacity should be and makes the use of the traditional replication argument in the pricing of options and futures questionable.

- Incentive Issues

Although we were able to obtain some preliminary results regarding the existence of a Nash equilibrium between buyers and sellers under the co-existence of spot markets and long-term contracts, the overall economics of B2B exchanges, including the role of suppliers, is still unclear. Also, at a more operational level, the creation of spot markets raises the issue of how existing (distribution) channels will react to this new channel. Dynamic game models can be used to model these environments.

- Collaborative Games

In principle, the aggregation of demand or supply can lead to significant benefits for the parties involved. Academic models could study the underlying dynamics and quantify the economics for the individual players.

- Fairness Issue

As stated above, the question of fairness in B2B exchanges represents a major challenge for B2B e-commerce. By the same token, it is not clear how the benefits due to exchanges, which are only the result of the interaction of several players, should be shared.

- Multi-criteria Decision Making

It is apparent that the corporate decision-makers are reluctant to embrace Internet-based marketplaces even if they offer lower procurement prices than established channels. Therefore, analytical models trading off several factors such as lead-time performance, quality of supply, and unit price, will be of great value to practitioners and will be part of our future work.

References

AMR Research, Inc. (2000) *Evaluating the Independent Trading Exchanges.*

Anupindi, R. and R. Akella (1993) "Diversification under Supply Uncertainty," *Management Science,* Vol. 39, No. 8, pp. 944 - 963

Araman, V., J. Kleinknecht, R. Akella, E. Tse, T. Costy, and J. Tew (2000a) "Supplier and Procurement Risk Management in E-Business: Optimal Long-Term and Spot Market Mix," Contract Report, GM Research & Development Center, Warren, MI

Araman, V., J. Kleinknecht, and R. Akella (2000b) "Supplier and Procurement Risk Management in E-Business," Working Paper, Department of Management Science & Engineering, Stanford University, Stanford, CA.

A.T. Kearney (2000) *E-pricing: How the Digital Marketplace is Changing Product Pricing*

Chopra S. and P. Meindl (2000) *Supply Chain Management: Strategy, Planning and Operation* Prentice Hall, NJ

Cohen, M.A. and N. Agrawal (1999) "An Analytical Comparison of Long and Short Term Contracts," *IIE Transaction,* Vol. 31, pp. 783-796

Barnes-Schuster, D., Y. Bassok, and R. Anupindi (1998) "Coordination and Flexibility in Supply Contracts with Options," Working Paper, Graduate School of Business, University of Chicago, Chicago, IL. *Submitted for Publication.*

Bassok, Y. and R. Anupindi (1997a) "Analysis of Supply Contracts with Total Minimum Commitment," *IIE Transactions,* Vol. 29, No. 5, pp. 373-381

Bassok, Y. and R. Anupindi (1997b) "Analysis of Supply Contracts with Commitments and Flexibility," Working Paper, University of Washington, Seattle, WA

Bensaou, M. (1999) "Portfolios of Buyer-Supplier Relationships," *Sloan Management Review,* Summer 1999, pp. 35-44

Durlacher (2000) *Business to Business E-Commerce*

Fong, D. K. H., V. M. Gempesaw, and J. K. Ord (2000) "Analysis of a Dual Sourcing Inventory Model with Normal Unit Demand and Erlang Mixture Lead Times," *European Journal of Operations Research,* Vol. 120, No. 1, pp. 97-107

Forbes ASAP (2000) "B2B ... to Be?" Special Section, August 21

Goldman Sachs (1999) *B2B: 2B or Not 2B?*

Kaplan, S. and M. Sawhney (2000) "E-Hubs: The new B2B Marketplaces," *Harvard Business Review,* May-June 2000, pp. 97-103

Kleinknecht, J., J. Weyant, and R. Akella (2001) "Optimal Supply Contract Portfolios in E-Business," Working Paper, Department of Management Science & Engineering, Stanford University, Stanford, CA.

Lee, H., V. Padmanabhan, and S. Whang (1997) "Information Distortion in a Supply Chain: The Bullwhip Effect," *Management Science,* Vol. 43, No. 4, pp. 546-558

Li, C. and P. Kouvelis (1999) "Flexible and Risk-Sharing Supply Contracts under Price Uncertainty," *Management Science,* Vol. 45, No. 10, pp. 1378-1398

Morgan Stanley Dean Witter (2000) *Collaborative Commerce*

Motley Fool Research (2000) "Industry Analysis: Business-to-Business E-Commerce," Internet Report, pp. 6 - 15

Rudi, N. (1999) "Dual Sourcing: Combining Make-to-Stock and Assemble-to-Order," Working Paper, The Simon School, University of Rochester, Rochester, NY.

Sawhney M. and S. Kaplan (1999) "Let's Get Vertical," Business 2.0, September, pp. 86-91

Seifert, R. and U. Thonemann (2000) "Optimal Procurement Strategies for Online Spot Markets," Working Paper, Department of Management Science and Engineering, Stanford University, Stanford, CA, *Submitted for Publication*

Tsay, A, S. Nahmias and N. Agrawal (1998) "Modeling Supply Chain Contracts: A Review," S. Tayur, R. Ganeshan, M. Magazine, eds., Quantitative Models for Supply Chain Management. Kluwer Academic Publishers, Boston, MA

II

SUPPLY CHAIN COORDINATION MODELS AND APPLICATIONS

Chapter 4

MANAGING DEMAND UNCERTAINTY FOR SHORT LIFE CYCLE PRODUCTS USING ADVANCE BOOKING DISCOUNT PROGRAMS

Christopher S. Tang
The Anderson School
UCLA
Los Angeles, CA 90095
chris.tang@anderson.ucla.edu

Kumar Rajaram
The Anderson School
UCLA
Los Angeles, CA 90095
kumar.rajaram@anderson.ucla.edu

Jihong Ou
Department of Decision Sciences
Faculty of Business Administration
National University of Singapore
Kent Ridge, Singapore
fbaoujh@nus.edu.sg

Abstract Consider a retailer that sells perishable seasonal products with uncertain demand. Due to the short sales season and long replenishment lead times associated with such products, the retailer is unable to update demand forecasts by using actual sales data generated from the early part of the season and to respond by replenishing stocks during the season. To overcome this limitation, we examine the case in which the retailer develops a new program called 'Advance Booking Discount'

J. Geunes et al. (eds.),
Supply Chain Management: Models, Applications, and Research Directions, 69–95.

(ABD) program that entices customers to pre-commit their orders at a discount price prior to the selling season. However, such orders are filled during the selling season. The time between the placement and the fulfillment of these pre-committed orders provides an opportunity for the retailer to update demand forecasts by utilizing information generated from the pre-committed orders and to respond by placing a cost effective order at the beginning of the selling season. In this chapter, we evaluate the benefits of the ABD program and characterize the optimal discount price that maximizes the retailer's expected profit.

1. Introduction

To compete in global markets, many companies continue to launch new products and phase out old products rapidly.[1] As product life cycles shorten, the fundamental issues in managing interacting areas such as pricing, forecasting and inventory control mimic those of fashion products. For instance, if a company over forecasts and orders more than the actual demand, then it has to reduce prices so as to sell the leftover inventory at the end of the selling season[2]. Responding accurately to changing demand patterns by forecasting and translating forecasts into an efficient supply plan are key ingredients for success, especially for products with short life cycles and high demand uncertainty.

Fisher and Raman (1996) discuss the strategy of 'accurate response' in which manufacturers first utilize early season sales data to update demand forecasts, and then respond to updated demand forecasts by producing and delivering products within the sales season. This strategy has helped Sport Obermeyer to make significant improvements in terms of inventory reduction, customer service and net profit. However, accurate response has two crucial requirements. First, the sales season has to be sufficiently long and fairly representative in the beginning of the season (so that one can update the demand forecast by using the sales data during the early part of the selling season). Second, the replenishment lead time has to be shorter than the selling season (so that one can respond to the updated demand forecast by replenishing stocks within the selling season). These requirements may not be met in various situations. For example, consider the sales of pumpkin pies at a supermarket during Thanksgiving. The selling season is very short (approximately 3 days), which makes it difficult to capture the sales data

[1]Billington, Lee and Tang (1998) report a list of challenging problems associated with rapid product replacements, which includes accurate demand forecasting and inventory management.

[2]Pashigan (1988) reports that average mark down for fashion merchandise in the apparel industry is around 16%.

during the early part of the selling season. Since the replenishment lead times are usually long (approximately 5 to 7 days), it is difficult for the bakery to respond to an updated demand forecast. Consequently, to make supply meet demand under these circumstances, one needs to consider other strategies.

In this chapter, we consider an alternative strategy under which the retailer develops an 'Advance Booking Discount' (ABD) program that entices customers to pre-commit their orders at a discount price prior to the selling season. However, these pre-committed orders are non-cancelable and are filled during the selling season. While the origin of the ABD program is unknown, we have observed its practice at Maxim's bakery in Hong Kong. Maxim's bakery, the largest bakery in Hong Kong, owns all of the cakes hops located at all subway stations in Hong Kong. Maxim's dominates sales in the baked goods market in Hong Kong largely due to its reputation for quality and convenience. Around 6 years ago, Maxim's launched the ABD program for the sales of moon cakes–a traditional Chinese cake composed of a stuffing made from lotus seed paste and egg yolk. The moon cake is a perishable seasonal food consumed by the Chinese when celebrating the mid-Autumn festival. Maxim's ABD program operates in the following manner: During the month prior to the mid-Autumn festival, customers can place their orders at any of the Maxim's cake shops at 25% off the regular price. Customers pay the discounted price when placing their orders in advance and receive redemption coupons for pick up during the week prior to the mid-Autumn festival. No order cancellation or refund is permitted. Maxim's guarantees the availability of the moon cakes only to those customers who participate in the ABD program. If customers do not participate in the ABD program, they can always try to buy the cake during the week prior to the mid-Autumn festival at the regular price.

There are three important benefits associated with the ABD program that will enhance the supply chain performance of the retailer. First, the ABD program extends the selling season without the need for immediate delivery. This enables the retailer to entice more customers to buy the product over a longer period of time without being constrained by production capacity. Second, under the ABD program, the placement of the pre-committed orders takes place prior to the season while the fulfillment of these pre-committed orders occurs during the season. Therefore, the time window between these two events provides an opportunity for the retailer to utilize advance booking data to generate a better demand forecast prior to the start of the selling season. Such improved forecasts offer an opportunity to the retailer for placing a more accurate order at the start of the season, while in turn reducing overstock and under-stock costs and improving customer service levels. Finally, the ABD program allows the retailer to improve cash flows because the payment from those advance bookings is received prior to the selling season.

In this chapter, we model the decisions under the ABD program, which involve how much to discount, how to use the pre-committed orders to update forecasts, and how much to order at the beginning of the season. In addition, our model allows us to explicitly quantify the first two benefits of the ABD program. We compare the profit associated with this program to that of the traditional sales (no early promotion) program. We also characterize the conditions under which the ABD program is beneficial to the retailer.

This chapter is organized as follows. Section 2 provides a brief review of marketing and operations management literature that deals with promotional discount. In section 3, we first present the base model (for the case with no discount promotion) and then we present the ABD model. We also compare the profit associated with the base case to that of the ABD model and characterize the conditions under which the ABD program is beneficial. Section 4 analyzes the properties of the optimal discount price and provides numerical examples to illustrate our basic results. We present two extensions in section 5, and end with concluding remarks in section 6.

2. Literature Review

Most of the research on quantity discounts can be classified into two streams. The first stream focuses on the analysis of the optimal ordering policy for a buyer when the demand is constant and the supplier offers a specific discount policy. The second stream examines how a supplier can use a discount policy as a control mechanism to induce a buyer to coordinate the channels of distribution. Weng (1995) presents a model that integrates these two streams of work. The reader is referred to Weng (1995) and the comprehensive references therein.

To our knowledge, Weng and Parlar (1999) is the first chapter that presents a model in which the retailer offers a price discount to induce customers to commit their purchases prior to the beginning of the selling season. They determine the optimal order quantity for the retailer and characterize the optimal discount rate. While our chapter addresses a similar problem, our model differs from their model in several aspects. First, their model deals with the case in which the customers belong to a single market segment while our model deals with two segments. We believe that the two-segment model allows us to capture heterogeneous consumer preferences. Second, they assume that the pre-committed orders, generated by the program, are deterministic while the remaining demand occurred during the season is stochastic. In our model, we consider a more realistic case in which both the pre-committed orders and the demand occurring during the selling season are stochastic. Third, we consider the case in which the retailer would utilize the pre-committed orders to update the probability distribution of the remaining demand during the season, while Weng and Parlar do not model the issue of forecast updating. We think demand forecast updating is of critical importance because updated demand forecasts allow the retailer to place a more accurate order at the beginning of the season. Finally, while Weng and Parlar focus on determining the optimal order quantity and discount rate, our emphasis is on examining the benefits of the ABD program. Specifically, we are interested in analyzing general conditions under which the ABD program is beneficial, and examining the impact of demand uncertainty and market share on the optimal discount factor. Our goal is to develop managerial insights for when such programs should be instituted.

3. The Analysis Framework

Consider a retailer that sells a seasonal product that belongs to Brand A. The unit cost, selling price and salvage value of Brand A are c, p, and s, respectively. There are two customer segments: one buys Brand A and the other buys Brand B.[3] The demand generated by the segment who buys Brand A, denoted by D_A, is assumed to be normally distributed with mean μ_A and standard deviation σ_A. Let θ be the coefficient of variation, where $\theta = \frac{\sigma_A}{\mu_A}$. Similarly, the demand generated by the segment who buys Brand B, denoted by D_B, is normally distributed with mean μ_B and standard deviation σ_B. To simplify the exposition, we assume that D_B has the same coefficient of variation so that $\theta = \frac{\sigma_B}{\mu_B}$; however, D_A and D_B are correlated with correlation coefficient $Corr(D_A, D_B)$. Let μ be the expected total market demand, where $\mu = \mu_A + \mu_B$. Suppose we set $\mu_A = \alpha\mu$ and $\mu_B = (1-\alpha)\mu$; then a can be interpreted as the market share of Brand A. By definition, $\sigma_A = \theta\alpha\mu$ and $\sigma_B = \theta(1-\alpha)\mu$.

3.1 The Base model

Consider the (base) case in which the retailer does not offer the ABD program. Thus, the retailer charges p for each unit during the season and charges s for each unit after the season. The retailer needs to determine

[3]Without loss of generality, Brand B is the aggregation of all other brands that compete with Brand A, and the retailer does not carry Brand B.

the optimal order quantity $Q*$ that maximizes the total expected profit. Let π be the optimal expected profit, where:

$$\pi = \text{Max}_Q \quad E_{D_A}\{p \min\{Q, D_A\} + s(Q - D_A)^+ - cQ\}.$$

The above problem is the newsvendor problem with normally distributed demand. It is well known that the optimal order quantity $Q^* = \mu_A + k\sigma_A$ and the optimal expected profit π is given as:

$$\pi = (p - c)\mu_A - (p - s)\phi(k)\sigma_A = \alpha\mu[(p - c) - (p - s)\phi(k)\theta], \quad (4.1)$$

where $k = \Phi^{-1}(\frac{p-c}{p-s})$, and $\Phi(.)$ and $\phi(.)$ are the distribution and the density functions of the standard normal distribution, respectively.

3.2 The Advance Booking Discount Model

Notice from (4.1) that the term $(p-s)\phi(k)\sigma_A = [(p-c)+(c-s)]\phi(k)\sigma_A$ corresponds to the sum of the expected overstock and understock costs associated with the optimal order quantity $Q*$. Thus, one can reduce the impact of these costs by reducing the demand variance σ_A^2. We now discuss how the ABD program can enable a retailer to achieve variance reduction. Under this program, the retailer offers a discount price xp per unit of Brand A (i.e., the discount factor is equal to x) prior to the beginning of the season, where $0 \leq x \leq 1$. If customers accept this offer, then they place an order by pre-paying xp per unit prior to the beginning of the season and pick up this order during the season. If customers decline this offer, they can always purchase the product during the season by paying regular price p per unit; however, the availability of the product will not be guaranteed.

The ABD program affects the two segments of customers as follows. First, among those customers who plan to buy Brand A during the selling season (i.e., D_A), $f(x)D_A$ will commit their orders at a lower price xp prior to the selling season and $(1 - f(x))D_A$ will purchase the product at regular price p during the selling season.[4] Second, for those customers who plan to buy Brand B during the selling season (i.e., D_B), $g(x)D_B$ will switch from buying Brand B to Brand A at a lower price xp prior to the selling season and the remaining $(1 - g(x))D_B$ will buy Brand B during the selling season as planned.[5] We assume that the

[4] It is conceivable that the consumption may increase as a result of price discount. However, since the product is perishable and since the customers can pick up the product only during the selling season, the customers may not be able to consume more during the selling season as a result of the ABD program. Thus, it seems reasonable to assume that the consumption remains the same.

[5] Since this segment plans to buy Brand B during the selling season, they would not buy Brand A during the selling season at the regular price p. However, some of them may switch

functions $f(x)$ and $g(x)$ are bounded between 0 and 1 and decreasing in x, so that more customers will buy Brand A prior to the selling season as x decreases.

Let $D_1(x)$ be the pre-committed orders occurred prior to the season and let $D_2(x)$ be the demand that occurs during the season, where:

$$\begin{aligned} D_1(x) &= f(x)D_A + g(x)D_B, \\ D_2(x) &= (1 - f(x))D_A. \end{aligned}$$

Notice that $D_1(x) + D_2(x) = D_A + g(x)D_B \geq D_A$, because the ABD program generates additional demand for Brand A due to customers who switch from buying Brand B at its regular price to Brand A at the discount price xp.

Suppose that the joint distribution of $D_1(x)$ and $D_2(x)$ is a bivariate normal distribution with means μ_1, and μ_2, standard deviations s_1 and s_2, and correlationcoefficient ρ, where

$$\begin{aligned} \mu_1 &= f(x)\mu_A + g(x)\mu_B = f(x)\alpha\mu + g(x)(1-\alpha)\mu, && (4.2) \\ \mu_2 &= (1 - f(x))\mu_A = (1 - f(x))\alpha\mu, && (4.3) \\ s_1 &= [f^2(x)\alpha^2\mu^2\theta^2 + g^2(x)(1-\alpha)^2\mu^2\theta^2 \\ &\qquad +2f(x)g(x)\alpha(1-\alpha)\mu^2\theta^2 Corr(D_A, D_B)]^{1/2}, && (4.4) \\ s_2 &= (1 - f(x))\alpha\mu\theta, && (4.5) \\ \rho &= \frac{cov((D_1(x), D_2(x))}{\sigma_{D_1(x)}\sigma_{D_2(x)}} \\ &= \frac{f(x) + g(x)rCorr(D_A, D_B)}{\sqrt{f^2(x) + g^2(x)r^2 + 2f(x)g(x)rCorr(D_A, D_B)}}. && (4.6) \end{aligned}$$

where $r = \frac{1-\alpha}{\alpha}$. Then it is well known (Bickel and Doksum (1977)) that the distribution of $D_2(x)$ given $D_1(x)$ (i.e., $(D_2(x)|D_1(x) = d_1)$) is also

to buy Brand A prior to the selling season because of the discount price associated with the ABD program.

normally distributed with mean μ' and standard deviation σ', where[6]:

$$\begin{aligned}
\mu' &= \mu_2 + \rho(d_1 - \mu_1)s_2/s_1, && (4.7)\\
\sigma' &= s_2\sqrt{1-\rho^2}\\
&= \sqrt{\frac{g^2(x)r^2(1-[Corr(D_A,D_B)]^2)}{f^2(x)+g^2(x)r^2+2f(x)g(x)rCorr(D_A,D_B)}} \times\\
&\quad (1-f(x))\alpha\mu\theta. && (4.8)
\end{aligned}$$

We offer two observations. First, notice from (4.8) that $\sigma' \le s_2$. Thus, the ABD program will reduce the variance of the demand that occurs during the selling season. Second, it can be shown from (4.8) that $(\sigma')^2$ is concave in $Corr(D_A, D_B)$ and that $(\sigma')^2$ is decreasing in $Corr(D_A, D_B)$ for $Corr(D_A, D_B) > 0$. Thus, positively correlated demands further reduce demand variance $(\sigma')^2$ due to the additional information associated with pre-committed orders generated from those customers switching from Brand B to Brand A. These two observations illustrate the basic mechanism by which the ABD program enables the retailer to increase sales, obtain an improved forecast, and place a more accurate order so as to achieve higher expected profits.

In the remainder of this section, we shall evaluate the optimal expected profits associated with the ABD program. To obtain tractable analytical results, we shall consider the case in which the correlation coefficient between D_A and D_B is equal to 0; (i.e., $Corr(D_A, D_B) = 0$). However, the same analysis can be extended numerically to include the case for which $Corr(D_A, D_B) \neq 0$.

When $Corr(D_A, D_B) = 0$, it is easy to check from (4.4), (4.6) and (4.8) that s_1, ρ and σ' can be expressed as follows:

$$\begin{aligned}
s_1 &= \sqrt{f^2(x)\alpha^2\mu^2\theta^2 + g^2(x)(1-\alpha)^2\mu^2\theta^2}\,, && (4.9)\\
\rho &= \frac{f(x)}{\sqrt{f^2(x)+g^2(x)r^2}}\,, && (4.10)\\
\sigma' &= \sqrt{\frac{g^2(x)r^2}{f^2(x)+g^2(x)r^2}}(1-f(x))\alpha\mu\theta. && (4.11)
\end{aligned}$$

[6]The bivariate normal distribution allows us to obtain simple expressions for μ' and σ' and to simplify our analysis. To elaborate, if one uses the conjugate prior distributions to determine the posterior distribution of the updated demand, then the mean and the standard deviation of the posterior distribution is quite complex and would complicate the analysis significantly. Also, for the case when the retailer offers no discount; i.e., when $x = 1$, we have $f(x) = g(x) = 0$, and $(D_1(x), D_2(x))$ has a degenerate bivariate normal distribution that has $\rho = 0$. In such case, we have $\mu' = \mu_2$ and $\sigma' = s_2$.

3.2.1 No Demand Forecast Updating. Consider the case when the retailer offers the ABD program with the discount factor x. To isolate the benefits of variance reduction and improved forecast due to updating in the ABD program, we first assume that the retailer is unable to utilize the pre-committed orders $D_1(x)$ to update the distribution of $D_2(x)$.[7] This scenario is plausible when the retailer lacks the infrastructure to capture or analyze sales data.

Since the order is placed at the start of selling season, the retailer can order the exact amount to fulfill the pre-committed orders $D_1(x)$ observed prior to the selling season. Hence, the profit generated from those pre-committed orders will equal $(xp - c)D_1(x)$. Although the retailer does not use $D_1(x)$ to update the distribution of $D_2(x)$, $D_2(x)$ is still normally distributed with mean μ_2 and standard deviation s_2 given by (4.3) and (4.5), respectively. In this case, the retailer orders an additional quantity Q' so as to cover the demand during the selling season. Thus, the profit generated from the demand $D_2(x)$ is equal to $\{p\min\{Q', D_2(x)\} + s(Q' - D_2(x))^+ - cQ'\}\}$. The total expected profit associated with the ABD program *without* demand forecast updating, denoted by $\hat{\pi}(x)$, can be expressed as follows:[8]

$$\begin{aligned}\hat{\pi}(x) &= E_{D_1(x)}\{(xp - c)D_1(x)\\ &\quad + \text{Max}_{Q'} E_{D_2(x)}\{p\min\{Q', D_2(x)\} + s(Q' - D_2(x))^+ - cQ'\}\} \,.\end{aligned}$$

By using the standard newsvendor result, the expected profit $\hat{\pi}(x)$ can be expressed as:

$$\begin{aligned}\hat{\pi}(x) &= (xp - c)(f(x)\alpha + g(x)(1 - \alpha))\mu + (p - c)(1 - f(x))\alpha\mu\\ &\quad - (p - s)\phi(k)(1 - f(x))\alpha\mu\theta. \qquad (4.12)\end{aligned}$$

We now compare the expected profit $\hat{\pi}(x)$ for any discount factor x given in (4.12) with the optimal profit π associated with the base case given in (4.1). When $x = 1$, $f(1) = g(1) = 0$, and $\hat{\pi}(1) = \pi$. Thus, the optimal expected profit $\text{Max}_{x\in[0,1]}\ \hat{\pi}(x)$ must be at least equal to π. In this case, we have proved the following Lemma:

Lemma 4.1 $Max_{x\in[0,1]}\ \hat{\pi}(x) \geq \pi$.

The above lemma implies that the ABD program, even without updated demand forecasts, can increase the retailer's expected profit due to two

[7]We shall consider the case in which the retailer would utilize the early sales information to update the demand forecast in section 3.2.2

[8]Without loss of generality, we omit the fixed promotion cost associated with the ABD program. Clearly, when the promotion cost is prevalent, the decision maker can always check to see if the expected benefit of the ABD program outweighs this promotion cost.

reasons. First, this program generates additional demand because it offers an economic incentive for customers to switch from Brand B to Brand A by paying a reduced price xp. Second, this program reduces the demand uncertainty because a portion of the demand has been made certain (i.e., the pre-committed orders) prior to the sales season.

3.2.2 Demand Forecast Updating. Consider the case in which the retailer offers the ABD program and utilizes the pre-committed orders $D_1(x)$ to update the distribution of $D_2(x)$. In this case, the profit generated from those pre-committed orders equals to $(xp - c)D_1(x)$. Since the updated distribution of $D_2(x)$ is $D_2(x)|D_1(x)$, the retailer would order additional quantity Q' so as to cover the demand during the season.[9] Thus, the profit generated during the season is equal to $\{p\min\{Q', D_2(x)\} + s(Q' - D_2(x))^+ - cQ'\}\}$. The optimal total expected profit associated with the ABD program with demand forecast updating, denoted by $\tilde{\pi}(x)$, can be expressed as:

$$\begin{aligned}\tilde{\pi}(x) &= E_{D_1(x)}\{(xp - c)D_1(x) + \mathrm{Max}_{Q'} E_{D_2(x)|D_1(x)}\{p\min\{Q', D_2(x)\} \\ &\quad + s(Q' - D_2(x))^+ - cQ'\}\}.\end{aligned}$$

Since $D_2(x)|D_1(x)$ is normally distributed with mean μ' and standard deviation σ', we can utilize (4.7), (4.11), and the newsvendor result to express $\tilde{\pi}(x)$ as:

$$\begin{aligned}\tilde{\pi}(x) &= (xp - c)(f(x)\alpha + g(x)(1 - \alpha))\mu + (p - c)(1 - f(x))\alpha\mu \\ &\quad - (p - s)\phi(k)\sqrt{\frac{g^2(x)r^2}{f^2(x) + g^2(x)r^2}}(1 - f(x))\alpha\mu\theta. \qquad (4.13)\end{aligned}$$

We now compare the expected profit $\tilde{\pi}(x)$ given in (4.13) with the expected profit $\hat{\pi}(x)$ given in (4.12). For any given discount factor x, it can be shown that $\tilde{\pi}(x) = \hat{\pi}(x) + (p - s)\phi(k)(1 - \sqrt{\frac{g^2(x)r^2}{f^2(x)+g^2(x)r^2}})(1 - f(x))\alpha\mu\theta \geq \hat{\pi}(x)$. Therefore, we have proved the following Lemma:

Lemma 4.2 $Max_{x\in[0,1]}\ \tilde{\pi}(x) \geq Max_{x\in[0,1]}\ \hat{\pi}(x) \geq \pi$.

The above lemma implies that, when implementing the ABD program, the retailer can realize higher expected profits if the retailer utilizes the pre-committed orders $D_1(x)$ to update the demand distribution $D_2(x)$. This is because the variance of the demand $D_2(x)$ will be further reduced due to updating. Overall, by increasing sales and reducing demand variance, the ABD program enables the retailer to increase the expected profit.

[9] The total order quantity is now equal to $D_1(x) + Q'$.

4. The Optimal Discount Factor

Lemmas 4.1 and 4.2 imply that the ABD program can increase expected profits. We now characterize the *optimal discount factors* $\hat{x}$ and $\tilde{x}$ that maximize the profit functions $\hat{\pi}(x)$ and $\tilde{\pi}(x)$, respectively. To obtain some structural results, let $f(x) = g(x)$ for $0 \leq x \leq 1$. This corresponds to the case in which both segments have identical response to the discount price. When $f(x) \neq g(x)$, we conduct our analysis numerically in section 5.

We restrict our attention to cases for which $f(x)$ belongs to a class of functions that satisfies the following properties: $f(x)$ is bounded between 0 and 1 and $f(x)$ is decreasing in x. In particular, we set $f(x) = 1 - x^f$, where $f > 0$. The choice of this general form of $f(x)$ is useful in capturing various types of market responses to the ABD program. For instance, when $f > 1$, $f(x)$ is decreasing and concave in x and is bounded between 0 and 1. Here, customers are eager to accept the ABD offer by pre-committing their orders prior to the season as a small discount induces a large fraction of customers to switch over. On the other hand, when $0 < f < 1$, $f(x)$ is decreasing and convex in x and is bounded between 0 and 1. In this case, customers are more reluctant to accept the ABD offer since only a large discount causes a large fraction of customers to switch over.

4.1 No Demand Forecast Updating

We now analyze the difference in profits between the base case given in (4.1) and the case of no demand forecast updating given in (4.12). This difference is defined by $\hat{\Delta}(x) = \hat{\pi}(x) - \pi$. Since the profit associated with the base case π is independent of x, finding $\hat{x}$ that maximizes the profit $\hat{\pi}(x)$ is equivalent to finding $\hat{x}$ that maximizes the function $\hat{\Delta}(x)$. Thus, it suffices to focus on $\hat{\Delta}(x)$.

Prior to presenting the properties of $\hat{x}$, we define a term $\hat{x}^c$ and a function $\hat{h}(x)$ that are useful in simplifying the exposition. Let:

$$\hat{x}^c = \frac{c + (p-c)\alpha - (p-s)\phi\alpha\theta}{p}, \qquad (4.14)$$

$$\hat{h}(x) = (xp - c) - (p-c)\alpha + (p-s)\phi\alpha\theta. \qquad (4.15)$$

For convenience, $\phi(k)$ is abbreviated by ϕ for the remainder of this chapter.

Lemma 4.3 *$\hat{x}^c$ and $\hat{h}(x)$ have the following properties:*

1 $\hat{x}^c \leq 1$.

2 $\hat{h}(x) < 0 \ \ \forall\, x < \hat{x}^c$, $\hat{h}(x) = 0$ *when* $x = \hat{x}^c$, *and* $\hat{h}(x) > 0 \ \ \forall\, x \in (\hat{x}^c, 1]$.

Proof: All proofs are given in the Appendix.
When $f(x) = g(x)$, (4.1), (4.12), and (4.15) imply that:

$$\hat{\Delta}(x) \;=\; \hat{\pi}(x) - \pi = \mu f(x)\hat{h}(x). \tag{4.16}$$

When $f(x) = 1 - x^f$, the first derivative of $\hat{\Delta}(x)$ can be expressed as:

$$\hat{\Delta}'(x) \;=\; -\mu f x^{f-1}\hat{h}(x) + \mu p(1 - x^f). \tag{4.17}$$

Define

$$\hat{\theta} \;=\; \frac{c + (p-c)\alpha}{(p-s)\phi\alpha}. \tag{4.18}$$

The following Propositions describe the properties of the optimal discount factor $\hat{x}$ that maximizes $\hat{\Delta}(x)$ when $f(x) = 1 - x^f$.

Proposition 4.4 *$\hat{x}$ has the following properties:*

1 *If $\theta \leq \hat{\theta}$, then $\hat{x}^c \geq 0$, $\hat{x} \in (\hat{x}^c, 1]$, and $\hat{x}$ satisfies the first order condition $\hat{\Delta}'(\hat{x}) = 0$.*

2 *If $\theta > \hat{\theta}$, then $\hat{x}^c < 0$, $\hat{x} \in (0, 1]$ and $\hat{x}$ satisfies the first order condition $\hat{\Delta}'(\hat{x}) = 0$.*

Proposition 4.5 *$\hat{x}$ has the following additional properties:*

1 *If $\theta \leq \hat{\theta}$, then $\hat{x}$ is decreasing in θ; $\hat{x}$ is increasing in a when $\theta \in [0, \frac{p-c}{(p-s)\phi}]$, and decreasing in a when $\theta \in (\frac{p-c}{(p-s)\phi}, \hat{\theta}]$.*

2 *If $\theta > \hat{\theta}$, then $\hat{x}$ has the following properties:*

   (a) *If $f \geq 1$, then $\hat{x}$ is decreasing in θ and α.*

   (b) *If $1 > f > 0$, then a threshold θ^a exists such that $\hat{x}$ is decreasing in θ when $\theta \leq \theta^a$ (and increasing in θ, otherwise). In addition, a threshold θ^b exists such that $\hat{x}$ is decreasing in α when $\theta \leq \theta^b$ (and increasing in α, otherwise).*

3 *If $f = 1$, then $\hat{x}$ can be expressed as follows:*

$$\hat{x} \;=\; 1 - \frac{(p-c)(1-\alpha) + (p-s)\phi\alpha\theta}{2p} \tag{4.19}$$

Propositions 4.4 and 4.5 lend themselves to the following interpretations. When the underlying demand is stable (i.e., when $\theta \leq \hat{\theta}$), the optimal discount factor satisfies the first order condition (i.e., by setting (4.17) to zero). In addition, the first statement of Proposition 4.5 implies that: (a) It is optimal to offer a lower discount price as the demand becomes more unstable; (i.e., $\hat{x}$ is decreasing in θ). This is because lowering the discount price will make a larger portion of the demand (i.e., the pre-committed orders) certain and reduce the variance of the demand occurred during the season; and (b) When θ is less than $\frac{p-c}{(p-s)\phi}$, it is optimal to offer a higher discounted price when the brand market share a is high. This is because when α is high, there is only a small gain in additional demand, which does not justify lowering the discount price. Conversely, consider the case when θ is greater than $\frac{p-c}{(p-s)\phi}$. In order to dampen demand variance, it is optimal to offer a lower discounted price even if α is high. We can interpret these results in a similar manner for the case when the underlying demand is unstable (i.e., when $\theta > \hat{\theta}$).

In summary, the effectiveness of the ABD program and the optimal discount factor depend on the demand stability, θ, and the brand market share, α. Thus, knowledge about the demand characteristics (represented by the coefficient of variation θ), consumer behavior (represented by the parameter f), and market conditions (represented by α) are all key aspects for determining the viability of the ABD program.

4.1.1 Illustrative Example. To better illustrate Propositions 4.4 and 4.5, we construct a numerical example for which the relevant data parameters are summarized in Table 4.1. Substituting these parameters into (4.14) and (4.18), we get $\hat{x}^c = 0.71$ and $\hat{\theta} = 5.5$. In our example, we have $\theta = 0.3 < 5.5 = \hat{\theta}$.

Since $\theta < \hat{\theta}$, the first statement in Proposition 4.4 implies that $\hat{x} \in (\hat{x}^c, 1] = (0.71, 1]$. By using (4.16), we can compute $\hat{\Delta}(x)$ for the cases $f = 0.5, 1, 2$ by varying x from 0 to 1. In addition, we can compute the optimal discount factor $\hat{x}$ by setting the right hand size of (4.17) to zero. The first statement of Proposition 4.4 (i.e., $\hat{x} \in (\hat{x}^c, 1] = (0.71, 1]$) is illustrated by Figure 4.1, where the optimal discount factor $\hat{x}$ equals 0.96, 0.85, 0.87 for the cases when $f = 0.5, 1, 2,$ respectively.

To examine the impact of θ on $\hat{x}$, we vary θ from 0.1 to 5 so that $\theta < \hat{\theta}$. Thus, the first statement in Proposition 4.5 implies that $\hat{x}$ is decreasing in θ. This result is illustrated in Figure 4.2.

Next, we examine the impact of α on $\hat{x}$. By using the parameters given in Table 4.1, it is easy to show that $\frac{p-c}{(p-s)\phi} = 1.83$. We consider two cases in which $\theta = 0.3$ and $\theta = 2$ so that $\theta \in [0, \frac{p-c}{(p-s)\phi}]$ for the former

PARAMETER	VALUE
Price (p)	100
Cost (c)	50
Salvage Value (s)	25
Market Demand: Mean (μ)	100
Market Demand: Coefficient of Variation (θ)	0.3
Market Share (α)	0.5

Table 4.1. Parameters for illustrative example

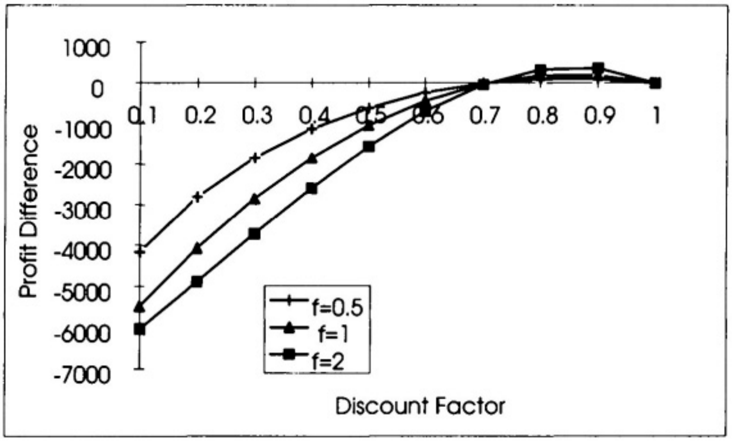


Figure 4.1. Profit Difference ($\hat{\Delta}(x)$) versus Discount Factor ($\hat{x}$) for ABD **without** updating

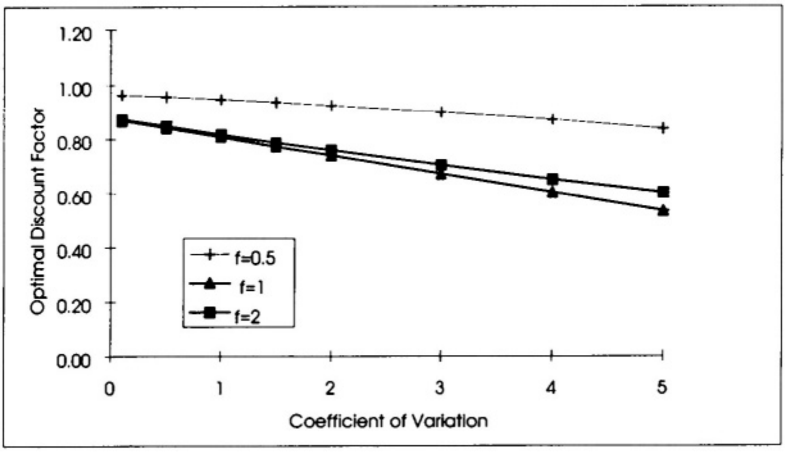


Figure 4.2. Optimal Discount Factor ($\hat{x}$) versus Coefficient of Variation (θ) for ABD **without** updating

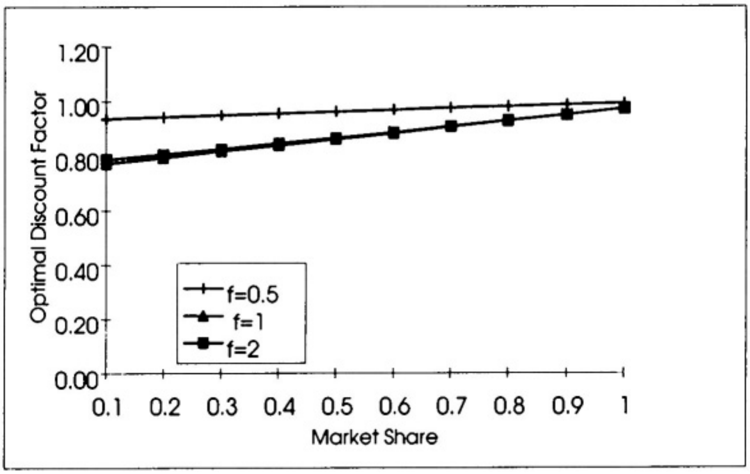


Figure 4.3. Optimal Discount Factor ($\hat{x}$) versus Market Share (α) when Coefficient of Variation (θ) = 0.3 for ABD **without** updating

case and $\theta \in (\frac{p-c}{(p-s)\phi}, \hat{\theta}]$ for the latter case. Since $\theta < \hat{\theta} = 5.5$ in both cases, the first statement in Proposition 4.5 implies that $\hat{x}$ is increasing in α for the former case and is decreasing in α for the latter case. These results are confirmed in Figures 4.3 and 4.4, respectively.

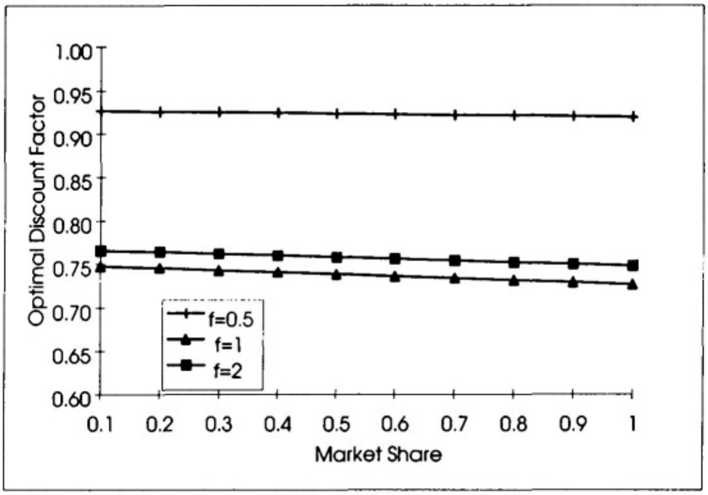


Figure 4.4. Optimal Discount Factor ($\hat{x}$) versus Market Share (α) when Coefficient of Variation (θ) = 2 for ABD **without** updating

4.2 Demand Forecast Updating

Let $\tilde{\Delta}(x) = \tilde{\pi}(x) - \pi$, where $\tilde{\Delta}(x)$ (analogous to $\hat{\Delta}(x)$) measures the profit difference between the ABD program with demand forecast updating and the base case. Since the base case is independent of x, it suffices to focus on $\tilde{\Delta}(x)$.

Prior to presenting the properties of the optimal discount factor $\tilde{x}$, we define two terms v and $\tilde{x}^c$, and a function $\tilde{h}(x)$ that are useful in simplifying the exposition. Let:

$$v = \sqrt{\frac{r^2}{1+r^2}} \leq 1, \text{ where } r = \frac{1-\alpha}{\alpha}, \tag{4.20}$$

$$\tilde{x}^c = \frac{c + (p-c)\alpha - (p-s)\phi v \alpha \theta}{p}, \tag{4.21}$$

$$\tilde{h}(x) = (xp - c) - (p-c)\alpha + (p-s)\phi v \alpha \theta. \tag{4.22}$$

Notice that the term $\tilde{x}^c$ and the function $\tilde{h}(x)$ are analogous to $\hat{x}^c$ and $\hat{h}(x)$ given in (4.14) and (4.15), respectively.

Lemma 4.6 *$\tilde{x}^c$ and $\tilde{h}(x)$ have the following properties:*

1 $\tilde{x}^c \leq 1$.

2 $\tilde{h}(x) < 0 \ \ \forall\, x < \tilde{x}^c$, $\tilde{h}(x) = 0$ *when* $x = \tilde{x}^c$, *and* $\tilde{h}(x) > 0 \ \ \forall\, x \in (\tilde{x}^c, 1]$.

When $f(x) = g(x)$, (4.1), (4.13), and (4.22) imply that:

$$\tilde{\Delta}(x) = \tilde{\pi}(x) - \pi = \mu f(x)\tilde{h}(x) + (p-s)\phi(1-v)\alpha\mu\theta. \quad (4.23)$$

When $f(x) = 1 - x^f$, the first derivative of $\tilde{\Delta}(x)$ can be expressed as:

$$\tilde{\Delta}'(x) = -\mu f x^{f-1}\tilde{h}(x) + \mu p(1 - x^f). \quad (4.24)$$

We define

$$\tilde{\theta} = \frac{c + (p-c)\alpha}{(p-s)\phi v\alpha}. \quad (4.25)$$

The following propositions describe the properties of the optimal discount factor $\tilde{x}$ that maximizes $\tilde{\Delta}(x)$ when $f(x) = 1 - x^f$.

Proposition 4.7 *$\tilde{x}$ has the following properties:*

1 *If $\theta \le \tilde{\theta}$, then $\tilde{x}^c \ge 0$, $\tilde{x} \in (\tilde{x}^c, 1]$, and $\tilde{x}$ satisfies the first order condition $\tilde{\Delta}'(\tilde{x}) = 0$.*

2 *If $\theta > \tilde{\theta}$, then $\tilde{x}^c < 0$, $\tilde{x} \in (0, 1]$ and $\tilde{x}$ satisfies the first order condition $\tilde{\Delta}'(\tilde{x}) = 0$.*

Proposition 4.8 *$\tilde{x}$ has the following additional properties:*

1 *If $\theta \le \tilde{\theta}$, then $\tilde{x}$ is decreasing in θ. In addition, if $(p-c) - (p-s)\phi\theta\frac{-1+r^3}{(1+r^2)^{1.5}} \ge 0$, then $\tilde{x}$ is increasing in α (and decreasing in α, otherwise).*

2 *If $\theta > \tilde{\theta}$, then the optimal discount factor $\tilde{x}$ has the following properties:*

   (a) *If $f \ge 1$, then $\tilde{x}$ is decreasing in θ. In addition, if $(p-c)-(p-s)\phi\theta\frac{-1+r^3}{(1+r^2)^{1.5}} \ge 0$, then $\tilde{x}$ is increasing in α (and decreasing in α, otherwise).*

   (b) *If $1 > f > 0$, then a threshold θ^d exists such that $\tilde{x}$ is decreasing in θ when $\theta \le \theta^d$ (and increasing in θ, otherwise). Moreover, a threshold θ^e exists such that $\tilde{x}$ is increasing in α when $\theta \le \theta^e$ and $(p-c) - (p-s)\phi\theta\frac{-1+r^3}{(1+r^2)^{1.5}} \ge 0$ or when $\theta > \theta^e$ and $(p-c) - (p-s)\phi\theta\frac{-1+r^3}{(1+r^2)^{1.5}} < 0$. Otherwise, $\tilde{x}$ is decreasing in α.*

3 *If $f = 1$, then $\tilde{x}$ can be expressed as follows:*

$$\tilde{x} = 1 - \frac{(p-c)(1-\alpha) + (p-s)\phi v\alpha\theta}{2p}. \quad (4.26)$$

Observe that Propositions 4.7 and 4.8 are analogous to Propositions 4.4 and 4.5 and thus, lend themselves to similar interpretations. Essentially, the demand stability θ and the brand market share α have a significant impact on the effectiveness and the optimal discount factor $\tilde{x}$ associated with the ABD program, and therefore, on the effectiveness of the ABD program.

The following Proposition compares the optimal discount factors that maximize the expected profits with and without demand forecast updating:

Proposition 4.9 *If $\theta \leq \tilde{\theta}$, then $\tilde{x} \geq \hat{x}$. Otherwise, $\tilde{x}$ has the following properties:*

1. *If $f \geq 1$, then $\tilde{x} \geq \hat{x}$.*
2. *If $1 > f > 0$, then a threshold θ^c exists such that $\tilde{x} \geq \hat{x}$ when $\theta \leq \theta^c$, and $\tilde{x} < \hat{x}$ otherwise.*

Proposition 4.9 and Lemma 4.2 have the following implication. If the retailer uses pre-committed orders to update the demand forecasts, then the retailer can achieve a higher expected profit by offering a higher price (i.e., $\tilde{x}p > \hat{x}p$). This is because the demand variance is further reduced when the retailer updates the demand distribution $D_2(x)$ after observing $D_1(x)$.

4.2.1 Illustrative Example. We now use the same numerical example presented in Table 4.1 to illustrate Propositions 4.7, 4.8, and 4.9. Substituting the parameters from Table 4.1 into (4.21) and (4.25), we get $\tilde{x}^c = 0.72$ and $\tilde{\theta} = 7.8$. In our example, we have $\theta = 0.3 < 7.8 = \tilde{\theta}$.

Since $\theta < \tilde{\theta}$, the first statement in Proposition 4.7 implies that $\tilde{x} \in (\tilde{x}^c, 1] = (0.72, 1]$. By using (4.23), we can compute $\tilde{\Delta}(x)$ for the cases $f = 0.5, 1, 2$ by varying x from 0 to 1. In addition, we can compute the optimal discount factor $\tilde{x}$ by solving (4.24). The first statement of Proposition 4.7 (i.e., $\tilde{x} \in (\tilde{x}^c, 1] = (0.72, 1]$) is illustrated by Figure 4.5, where the optimal discount factor $\tilde{x}$ equals 0.96, 0.86, 0.87 for the cases when $f = 0.5, 1, 2$, respectively.

To examine the impact of θ on $\tilde{x}$, we vary θ from 0.1 to 5 so that $\theta \leq \tilde{\theta}$. Thus, the first statement in Proposition 4.8 implies that $\tilde{x}$ is decreasing in θ. This result is illustrated in Figure 4.6.

To examine the impact of α on $\tilde{x}$, let us consider two cases in which $\theta = 0.3$ and $\theta = 2$. When $\theta = 0.3$, $(p-c)-(p-s)\phi\theta\frac{-1+r^3}{(1+r^2)^{1.5}} \geq 0$ for $0 \leq \alpha \leq 1$. Combining this observation with the fact that $\theta = 0.3 < 7.8 = \tilde{\theta}$, the first statement in Proposition 4.8 implies that $\tilde{x}$ is increasing in α. This result

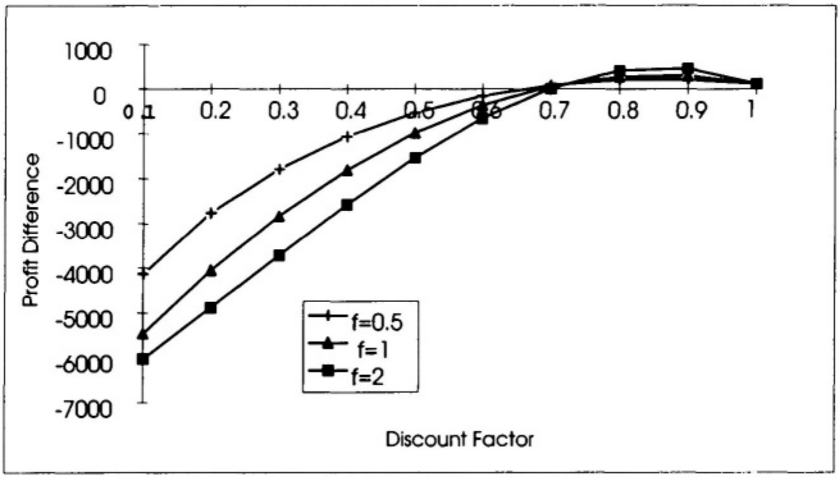


Figure 4.5. Profit Difference ($\tilde{\Delta}(x)$) versus Discount Factor ($\tilde{x}$) for ABD with updating

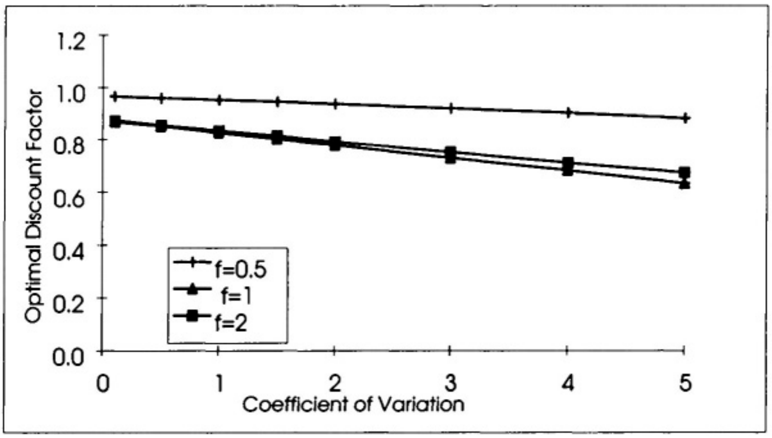


Figure 4.6. Optimal Discount Factor ($\tilde{x}$) versus Coefficient of Variation (θ) for ABD with updating

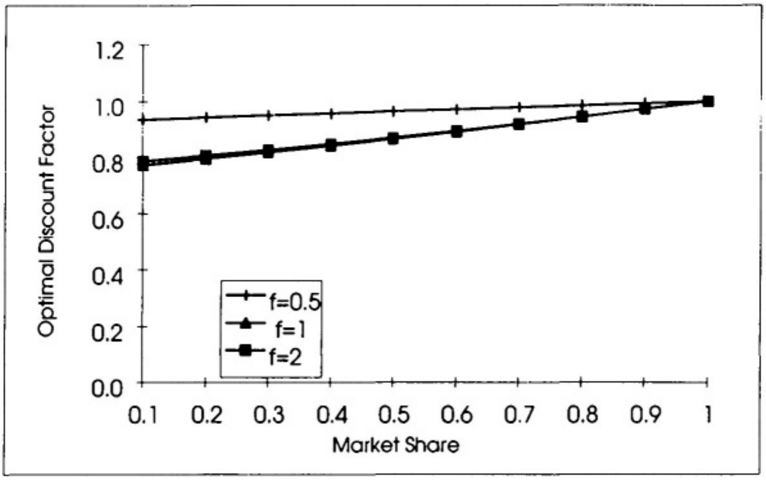


Figure 4.7. Optimal Discount Factor ($\tilde{x}$) versus Market Share (α) when Coefficient of Variation (θ) = 0.3 for ABD with updating

is illustrated in Figure 4.7. When $\theta = 2$, $(p-c)-(p-s)\phi\theta\frac{-1+r^3}{(1+r^2)^{1.5}} < 0$ for $0 \le \alpha < 0.5$, and $(p-c)-(p-s)\phi\theta\frac{-1+r^3}{(1+r^2)^{1.5}} \ge 0$ for $0.5 \le \alpha \le 1$. In this case, the first statement in Proposition 4.8 implies that $\tilde{x}$ is decreasing in α when $\alpha < 0.5$ and increasing in α when $\alpha \ge 0.5$. This result is illustrated in Figure 4.8.

Finally, we compare the optimal discount factor $\hat{x}$ for the no-updating case (Figure 4.2) and the optimal discount factor $\tilde{x}$ for the updating case (Figure 4.6). For the case when $\theta \le 5$, we found that $\tilde{x} \ge \hat{x}$ for the cases when $f = 0.5, 1$, and 2. Thus, this numerical result corroborates the first statement in Proposition 4.9.[10]

5. Non-identical Market Response Functions

We now numerically analyze the case when $f(x) = 1 - x^f$, $g(x) = 1 - x^g$, and $f \neq g$. As noted before, when $f > g > 1$, customers for Brand A are more eager to accept the ABD offer than customers for Brand B. The converse holds when $f < g$. We can also interpret the case when $g < f < 1$ and $f < g < 1$ in a similar manner. Throughout this section, we shall use the same numerical example presented in Table

[10]According to the second statement in Proposition 4.9, it is possible to have $\tilde{x} < \hat{x}$ when $f < 1$ and $\theta > \theta^c$. However, it can be shown that the value of $\theta^c \ge 100$ so that it is unlikely for this situation to occur in any real application.

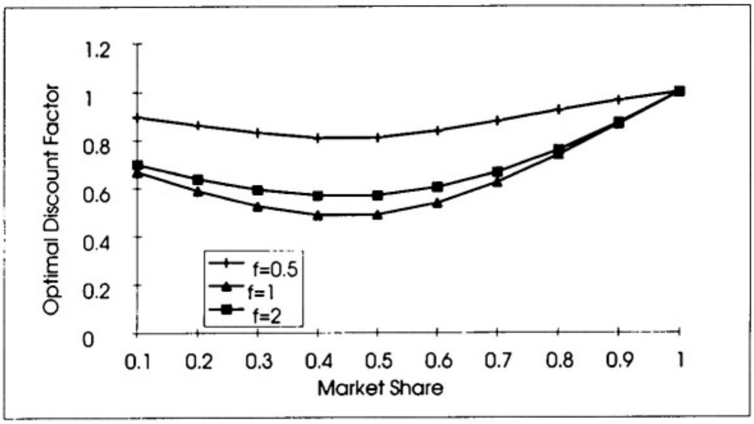


Figure 4.8. Optimal Discount Factor ($\tilde{x}$) versus Market Share (α) when Coefficient of Variation (θ) = 2 for ABD with updating

4.1, except that the parameter f is fixed at 1, and the parameter g is set to 0.5, 1, and 2, respectively.

5.1 Uncorrelated Demands

When the demands associated with segments A and B are uncorrelated (i.e., $Corr(D_A, D_B) = 0$), we can substitute $f(x)$ and $g(x)$ into (4.12) and (4.13) to show that:

$$\begin{aligned}
\hat{\pi}(x) &= (xp-c)((1-x^f)\alpha + (1-x^g)(1-\alpha))\mu \\
&\quad +(p-c)x^f\alpha\mu - (p-s)\phi(k)x^f\alpha\mu\theta\ , \qquad (4.27)\\
\tilde{\pi}(x) &= (xp-c)((1-x^f)\alpha + (1-x^g)(1-\alpha))\mu + (p-c)x^f\alpha\mu \\
&\quad -(p-s)\phi(k)\sqrt{\frac{(1-x^g)^2r^2}{(1-x^f)^2+(1-x^g)^2r^2}}(x^f\alpha\mu\theta). \qquad (4.28)
\end{aligned}$$

For any given value of θ, we compute the optimal discount factor $\tilde{x}$ that maximizes the expected profit function $\tilde{\pi}(x)$ for the case with demand forecast updating. Figure 4.9 summarizes our numerical results when we vary θ from 0.1 to 5.[11]

Figure 4.9 shows that $\tilde{x}$ is decreasing in θ, consistent with Figures 4.2 and 4.6 (for the case when $f = g$). In addition, it corroborates the first

[11] We also compute the optimal discount factor $\hat{x}$ that maximizes the expected profit function $\hat{\pi}(x)$ given in (4.27) for the case without demand forecast updating. However, since the numerical result has the same pattern as depicted in Figure 4.9, we omit the figure.

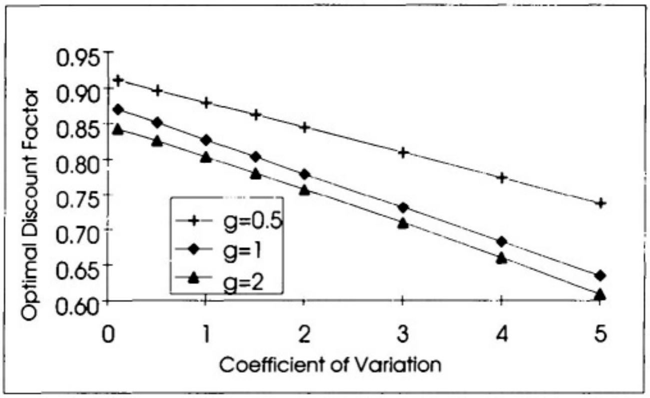


Figure 4.9. Optimal Discount Factor ($\tilde{x}$) versus Coefficient of Variation (θ) for ABD with updating when $f = 1$

statement of Propositions 4.5 and 4.8. Essentially, Figure 4.9 reports that, as demand becomes unstable (i.e., as θ increases), it is optimal for the retailer to offer a lower discounted price to induce more customers to pre-commit their orders prior to the selling season. As a larger portion of the demand now becomes certain, the variance of the remaining demand during the season is reduced.

In addition, Figure 4.9 illustrates the impact of different values of g on $\tilde{x}$. For a given value of θ, Figure 4.9 shows that $\tilde{x}$ decreases as g increases, as the customers for Brand B become more responsive toward the ABD program. This result seems to imply that as customers for Brand B become more eager to accept the offer associated with the ABD program, it is beneficial for the retailer to offer a lower discounted price to induce more customers to switch to Brand A.

5.2 Correlated Demands

Let the demands associated with segments A and B be correlated so that $Corr(D_A, D_B) \neq 0$. When the retailer does not utilize $D_1(x)$ to update the distribution of $D_2(x)$, the standard deviation of $D_2(x)$ during the season is equal to s_2 given in (4.5). Notice from (4.5) that s_2 is independent of $Corr(D_A, D_B)$. Consequently, under this situation, the retailer operates in the same manner, regardless of the correlation. For this reason, it suffices to focus only on the case with demand forecast updating. Suppose we substitute $f(x)$ and $g(x)$ into (4.8) and apply the same approach described in Section 3. Then we have:

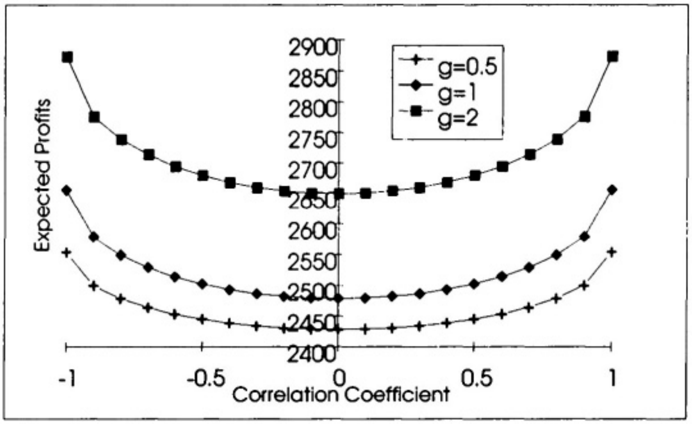


Figure 4.10. Optimal Expected Profits ($\tilde{\pi}(x)$) versus Correlation Coefficient ($\mathrm{Corr}(D_A, D_B)$) for ABD with updating when $f = 1$

$$\tilde{\pi}(x) = (xp - c)((1 - x^f)\alpha + (1 - x^g)(1 - \alpha))\mu + (p - c)x^f\alpha\mu$$
$$- (p - s)\phi(k)(x^f\alpha\mu\theta) \times$$
$$\sqrt{\frac{(1 - x^g)^2r^2(1 - [Corr(D_A, D_B)]^2)}{(1 - x^f)^2 + (1 - x^g)^2r^2 + 2(1 - x^f)(1 - x^g)rCorr(D_A, D_B)}}.$$

Figure 4.10 summarizes the optimal expected profit $\tilde{\pi}(x)$, while Figure 4.11 reports the optimal discount factor $\tilde{x}$ when we fix $\theta = 0.3$ and vary $Corr(D_A, D_B)$ from -1 to 1.

Figure 4.10 implies that, as the demands D_A and D_B become more (positively or negatively) correlated, the retailer can obtain a higher expected profit. This observation can be explained as follows. Under correlated demands, pre-committed orders generated from customers switching from Brand B to Brand A now provides informational value for predicting the remaining demand during the selling season. As described in Section 3, this additional information further reduces demand variance, which increases expected profits. Figure 4.11 suggests that it is optimal for the retailer to offer a lower discount when the demands D_A and D_B become more correlated. This result can be explained by the fact that under demand forecast updating, highly correlated demands naturally lead to a lower variance of forecast error. Therefore, it reduces the need to reduce forecast error through price discounting. Finally, let us examine how the optimal expected profit and the optimal discount factor are affected by the parameter g. Figure 4.10 implies that the retailer can achieve a higher expected profit as g increases (i.e., as the

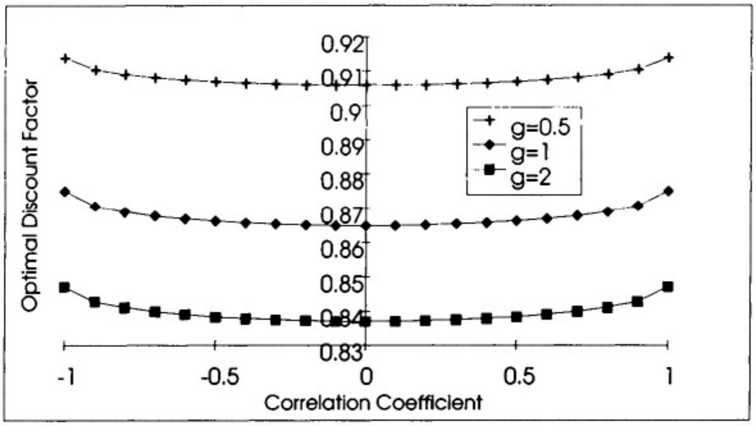


Figure 4.11. Optimal Discount Factor($\tilde{x}$) versus Correlation Coefficient (Corr(D_A, D_B)) for ABD with updating when $f = 1$

customers for Brand B become more responsive). This is because more customers for Brand B switch over to Brand A under the ABD program. Also, Figure 4.11 suggests that as g increases, it is beneficial for the retailer to offer a lower discounted price.

6. Concluding Remarks

In this chapter, we have considered a problem of matching supply with demand for products with short life-cycles and highly unpredictable demands. Due to long replenishment lead-time and a short sales season, the retailer is unable to re-stock during the selling season and respond to market demand. As an alternative strategy, we have considered a scheme called the Advance Booking Discount (ABD) program in which customers can pre-commit their orders, with guaranteed delivery during the season, at a discount price prior to the commencement of the sales season.

We have developed a model that enables us to quantify two crucial benefits of the ABD program including generating additional demand and better matching of supply with demand through more accurate forecasting and supply planning. Our analysis provides objective guidelines for deciding if, when and how such programs should be instituted. It is important to note that the ABD program presented in this chapter is amenable to many practical extensions including addition of a fixed cost for administering this program or the case with multiple products with fixed capacity constraints. Our future research focuses on testing the robustness of this model and results with alternate demand distributions and different types of customer response functions.

Acknowledgments

This research is partially supported by UCLA James Peters Research Fellowship. Ou participated in this project during a visit to UCLA in 1998.

Appendix

Proof of Lemma 2 Notice from (4.14) that $\hat{x}^c > 1$ if and only if $\theta < -\frac{(p-c)(1-\alpha)}{(p-s)\phi\alpha}$. This is impossible because the coefficient of variation $\theta \geq 0$. Thus, $\hat{x}^c \leq 1$. Next, observe from (4.15) that $\hat{h}(x)$ is increasing in x and $\hat{h}(\hat{x}^c) = 0$. The second statement of the Lemma follows immediately from these two observations.

Proof of Proposition 4.4 Since $\theta \leq \hat{\theta} = \frac{c+(p-c)\alpha}{(p-s)\phi\alpha}$, it is easy to check from (4.14) that $0 \leq \hat{x}^c \leq 1$. For any x that has $0 \leq x \leq \hat{x}^c \leq 1$, Lemma 2 implies that $\hat{h}(x) \leq 0$. In this case, one can show from (4.17) that $\hat{\Delta}'(x) \geq 0 \quad \forall x \in [0, \hat{x}^c]$. Next, since $\hat{x}^c \leq 1$, Lemma 2 implies that $\hat{h}(1) \geq 0$ and that $\hat{\Delta}'(1) \leq 0$. It follows from the Mean Value Theorem that there exists an optimal $\hat{x} \in (\hat{x}^c, 1]$ that satisfies the first order condition.

Next, when $\theta > \hat{\theta} = \frac{c+(p-c)\alpha}{(p-s)\phi\alpha}$, it is easy to check from (4.14) that $\hat{x}^c < 0$. In this case, we have $\hat{h}(x) > 0 \quad \forall x \in [0, 1]$. Hence, it can be shown from (4.17) that $\hat{\Delta}'(0) > 0$ and that $\hat{\Delta}'(1) < 0$. It follows from the Mean Value Theorem that there exists an optimal $\hat{x} \in (0, 1]$ that satisfies the first order condition. This completes the proof.

Before we present the proof of Proposition 4.5, let us prove the following Lemma that will be useful.

Lemma A.1 *Consider the function $\hat{h}(x)$ given in (4.15) and the function $f(x) = 1 - x^f$, where $f > 0$. The term $\hat{x}$, that has $\hat{\Delta}'(\hat{x}) = 0$, has the following properties:*

1. *When $\theta \leq \hat{\theta}$, the term $(f-1)\hat{h}(\hat{x}) + 2\hat{x}p \geq 0$.*

2. *When $\theta > \hat{\theta}$, the term $(f-1)\hat{h}(\hat{x}) + 2\hat{x}p \geq 0$ when $f \geq 1$. In addition, when $1 > f > 0$, the term $(f-1)\hat{h}(\hat{x}) + 2\hat{x}p \geq 0$ when θ is sufficiently small and the term $(f-1)\hat{h}(\hat{x}) + 2\hat{x}p < 0$ when θ is sufficiently large.*

Proof of Lemma A.1 Observe from (4.15) and (4.14) that $\hat{h}(x) = xp - \hat{x}^c p$. Hence, $(f-1)\hat{h}(\hat{x}) + 2\hat{x}p = (f+1)\hat{x}p - (f-1)\hat{x}^c p$. First, when $\theta \leq \hat{\theta}$, the first statement of Proposition 4.4 implies that $\hat{x} > \hat{x}^c \geq 0$. The term $(f-1)\hat{h}(\hat{x}) + 2\hat{x}p = (f+1)\hat{x}p - (f-1)\hat{x}^c p$ is non-negative because $\hat{x} > \hat{x}^c \geq 0$. Next, when $\theta > \hat{\theta}$, these cond statement of Proposition 4.4 implies that $\hat{x}^c < 0$ and $\hat{x} > \hat{x}^c$. In this case, the term $(f+1)\hat{x}p - (f-1)\hat{x}^c p$ is non-negative when $f \geq 1$. We finally consider the case when $1 > f > 0$. By substituting the expression for $\hat{x}^c$ given in (4.14) into the term $(f+1)\hat{x}p - (f-1)\hat{x}^c p$. It can be shown that this term is non-negative (negative) when θ is sufficiently small (large).

Proof of Proposition 4.5 Let us consider the case in which $\theta \leq \hat{\theta} = \frac{c+(p-c)\alpha}{(p-s)\phi\alpha}$.

First, considering the fact that $\hat{x}$ satisfies $\hat{\Delta}'(\hat{x}) = 0$, we can use the implicit function theorem to differentiate the function $\hat{\Delta}'(\hat{x})$ with respect to θ. By considering (4.17) and (4.15), it can be shown that:

$$\frac{d\hat{x}}{d\theta} = \frac{-\hat{x}(p-s)\phi\alpha}{(f-1)\hat{h}(\hat{x}) + 2\hat{x}p}.$$

By applying the result from Lemma A.1 to the above expression, one can see that $\frac{d\hat{x}}{d\theta} < 0$. Thus, $\hat{x}$ is decreasing in θ.

Next, let us differentiate the function $\hat{\Delta}'(\hat{x}) = 0$ with respect to α. By considering (4.17) and (4.15), it can be shown that:

$$\frac{d\hat{x}}{d\alpha} = \frac{\hat{x}[(p-c)-(p-s)\phi\theta]}{(f-1)\hat{h}(\hat{x}) + 2\hat{x}p}.$$

Since $\theta \leq \hat{\theta} = \frac{c+(p-c)\alpha}{(p-s)\phi\alpha}$, Lemma A.1 implies that the denominator of the above equation is non-negative. In this case, $\frac{d\hat{x}}{d\alpha} > 0$ if and only if the numerator $\hat{x}[(p-c)-(p-s)\phi\theta]$ is positive. The numerator is positive when $\theta < \frac{(p-c)}{(p-s)\phi}$. It follows from the definition of $\hat{\theta}$ that $\frac{(p-c)}{(p-s)\phi} < \hat{\theta}$. Therefore, we can conclude that the numerator is positive when $\theta \in [0, \frac{(p-c)}{(p-s)\phi})$ and is negative when $\theta \in [\frac{(p-c)}{(p-s)\phi}, \hat{\theta}]$. This proves the first statement.

For the case when $\theta > \hat{\theta}$, we can apply the same approach to prove the second statement. We omit the details.

Finally, when $f = 1$, we can determine the expression for $\hat{x}$ by setting $\hat{\Delta}'(\hat{x}) = 0$. We omit the details. This completes the proof.

Proof of Lemma 4.6 Since the term $\tilde{x}^c$ and the function $\tilde{h}(x)$ are analogous to $\hat{x}^c$ and $\hat{h}(x)$, we can use the same approach presented in the proof of Lemma 2 to prove Lemma 4.6. We omit the details.

Proof of Proposition 4.7 Observe that the term $\tilde{\theta}$ and the function $\tilde{h}(x)$ are analogous to the terms $\hat{\theta}$ and $\hat{h}(x)$. Also, notice that the first derivative of $\tilde{\Delta}(x)$ given in (4.24) is analogous to the first derivative of $\hat{\Delta}(x)$ given in (4.17). In this case, we can use the same approach presented in the proof of Proposition 4.4 to prove Proposition 4.7. We omit the details.

Before we present the proof of Proposition 4.8, let us prove the following Lemma that will be useful.

Lemma A.2 *Consider the function $\tilde{h}(x)$ given in (4.15) and the function $f(x) = 1 - x^f$, where $f > 0$. The term $\tilde{x}$, that has $\tilde{\Delta}'(\tilde{x}) = 0$, has the following properties:*

1. *When $\theta \leq \tilde{\theta}$, the term $(f-1)\tilde{h}(\tilde{x}) + 2\tilde{x}p \geq 0$.*
2. *When $\theta > \tilde{\theta}$, the term $(f-1)\tilde{h}(\tilde{x}) + 2\tilde{x}p \geq 0$ when $f \geq 1$. In addition, when $1 > f > 0$, the term $(f-1)\tilde{h}(\tilde{x}) + 2\tilde{x}p \geq 0$ when θ is sufficiently small and the term $(f-1)\tilde{h}(\tilde{x}) + 2\tilde{x}p < 0$ when θ is sufficiently large.*

Proof of Lemma A.2 Since the term $\tilde{x}^c$ and the function $\tilde{h}(x)$ are analogous to $\hat{x}^c$ and $\hat{h}(x)$, we can use the same approach presented in the proof of Lemma A.1 to prove Lemma A.2.We omit the details.

Proof of Proposition 4.8 First, differentiate the function $\tilde{\Delta}'(\tilde{x}) = 0$ with respect to θ and α, getting:

$$\frac{d\tilde{x}}{d\theta} = \frac{-\tilde{x}(p-s)\phi v\alpha}{(f-1)\tilde{h}(\tilde{x}) + 2\tilde{x}p},$$

$$\frac{d\tilde{x}}{d\alpha} = \frac{\tilde{x}[(p-c)-(p-s)\phi\theta\frac{dv\alpha}{d\alpha}]}{(f-1)\tilde{h}(\tilde{x}) + 2\tilde{x}p} = \frac{\tilde{x}[(p-c)-(p-s)\phi\theta\frac{-1+r^3}{(1+r^2)^{1.5}}]}{(f-1)\tilde{h}(\tilde{x}) + 2\tilde{x}p}$$

Then we can prove Proposition 4.8 by using the same approach as presented in the proof for Proposition 4.5. We omit the details.

Proof of Proposition 4.9 Let us compare the function $\tilde{\Delta}'(x)$ given in (4.24) and the function $\hat{\Delta}'(x)$ given in (4.17). It is easy to check from (4.22) and (4.15) that the key difference between $\tilde{\Delta}'(x)$ and $\hat{\Delta}'(x)$ lies with the term v given in (4.20). Essentially, $\tilde{\Delta}'(x)$ reduces to $\hat{\Delta}'(x)$ when $v = 1$. Since $v \leq 1$, we can prove the Proposition by showing that $\tilde{x}$ is decreasing in v. Let us differentiate the function $\tilde{\Delta}'(\tilde{x}) = 0$ with respect to v. By considering (4.24) and (4.22), it can be shown that :

$$\frac{d\tilde{x}}{dv} = -\frac{\tilde{x}[(p-s)\phi\alpha\theta]}{(f-1)\tilde{h}(\tilde{x}) + 2\tilde{x}p}.$$

By applying the result from Lemma A.2 to the above expression, we prove the Proposition. This completes the proof.

References

Bickel, P. and K. Doksum, *Mathematical Statistics,* Holden Day Publisher, San Francisco, 1977.

Billington, C., H. Lee, and C. Tang, "Product Rollovers: Process, Strategy," *Sloan Management Review,* vol. 39, pp. 23-30, (1998).

Fisher, M., and A. Raman, "Reducing the Cost of Demand Uncertainty through Accurate Response to Early Sales," *Operations Research,* vol. 44, pp. 87-99, (1996).

Pashigan, P., "Demand Uncertainty and Sales: A Study of Fashion and Markdown Pricing," *American Economic Review,* vol. 78, pp. 936-953, (1988).

Weng, K., and M. Parlar, "Integrating Early Sales with Production Decisions: Analysis and Insights," forthcoming, *IIE Transactions,* (1999).

Weng, K., "Channel Coordination and Quantity Discounts," *Management Science,* vol. 41, pp. 1509-1522, (1995).

Chapter 5

PARTIAL QUICK RESPONSE POLICIES IN A SUPPLY CHAIN

Craig E. Smith
McKinsey & Company, Inc.
1301 East 9th St.
Cleveland, OH 44114
Craig_Smith@McKinsey.com

Stephen M. Gilbert
Management Department
The University of Texas at Austin
CBA 4.202
Austin, TX 78712
steve.gilbert@bus.utexas.edu

Apostolos N. Burnetas
Department of Operations
Weatherhead School of Management
Case Western Reserve University
10900 Euclid Avenue
Cleveland, OH 44106
atb4@po.cwru.edu

Abstract It has been well documented that buyers can benefit significantly from being able to place reactive orders in response to observed demand for a short life cycle product. In practice, suppliers often fill these reactive orders with less than total reliability. Although reactive order fulfillment can allow the supply chain to capture more of the demand that is realized, it can also deter retailers from ordering as much initially. In this chapter, we investigate how this trade-off affects the retailers' ordering behavior as well as the profits of the manufacturer, the retailers,

J. Geunes et al. (eds.),
Supply Chain Management: Models, Applications, and Research Directions, 97–115.

and the supply chain as a whole. We also develop insight as to how a manufacturer should offer a reactive ordering policy.

1. Introduction and Related Literature

In many industries that are characterized by short product life cycles, manufacturers traditionally encourage retailers to order products well in advance of the selling season. In many cases, the entire season's demand has to be satisfied with one preseason order. For example, apparel retailers are often required to order products 3-6 months prior to the selling season, as discussed in Hammond and Raman, 1996 and Fisher and Raman, 1996. Often this is driven by manufacturers' willingness to forego responsiveness (i.e., short lead times) in return for low unit costs.

Recently, some manufacturers have begun to recognize the benefits of quick response systems, in which retailers are able to place and receive orders for additional quantities from the manufacturer during the selling season. This allows a retailer to adjust his quantity decision based on observations of early season sales. However, for a manufacturer to provide such responsiveness, she typically must either overproduce during a single production run or employ more expensive methods of production in order to produce on short notice. A considerable amount of analysis has been done to study precisely this trade-off. See, for example: Eppen and Iyer, 1997, Iyer and Bergen, 1997 and Lau and Lau, 1997.

In practice, manufacturers often provide the possibility, but not a guarantee, of quick response (Signorelli and Heskett, 1984). That is, when a retailer places an order after the start of the selling season, the manufacturer will fill it if she can, but does not guarantee that it will be filled. Clearly, a retailer's initial, preseason order will be affected by his confidence that the manufacturer will fill a subsequent order that he may place. If the manufacturer deals repeatedly with the same set of retailers, her history of order fulfillment influences the retailers' confidence that their orders will be filled. In this chapter, we investigate the way in which a manufacturer's quick response performance influences the size of the orders that are placed by retailers, and the profits of both members of the supply chain. In addition, we investigate the portion of all reorders that the manufacturer should optimally fulfill, and analyze the combination of wholesale price for reorders and portion of reorders filled.

The inclusion of a reorder opportunity introduces an interesting dynamic to the interaction between a manufacturer and her downstream retailers. By filling reorders, the manufacturer may be able to capture demand that would otherwise have been lost. At the same time, the greater the retailer's confidence that the reorder will be filled, the less he will tend to order initially. This will improve the profits of the supply chain as long as the benefit of ordering with better information offsets the potentially higher production, delivery, and backlog costs.

The remainder of the chapter is organized as follows. In section 2, we develop a model of a supply chain consisting of a manufacturer and a set of independent retailers in which the manufacturer provides partial fulfillment of reorders during the selling season. We analyze the model from the perspective of the manufacturer to determine how she should determine the portion of reorders to fill as well as the mark-up on the wholesale price for reorders. Our analytical results indicate that for uniform and exponential demand distributions, the manufacturer should provide either complete fulfillment of reorders, or no fulfillment whatsoever. Intermediate levels of fulfillment are never optimal. In Section 3, we perform numerical analysis to explore the effect of reorder fulfillment policies on channel profits and coordination. In addition, we investigate a variation of the original case, where the manufacturer makes only a single production run after receiving the retailers' initial orders, but can build inventory in anticipation of reorders. Finally, in Section 4, we discuss the practical implications of our results and suggest directions for future research. Throughout the chapter, we adopt the convention of using female pronouns to refer to the manufacturer, and using male pronouns to refer to the retailers.

2. Partial Fulfillment Model

Consider a setting in which a manufacturer sells her product through a set of *N* independent (i.e. non-competing) retailers. The manufacturer has two modes of production: one which is relatively inexpensive, but has a lead time sufficiently long that production quantities must be committed prior to the selling season; and the other which is more expensive but allows production to be done during the selling season. We denote the per-unit production costs of these two modes by c_1 and c_2 respectively.

The retailers have two opportunities to order the product: before and after observing demand. However, the manufacturer does not guarantee that the reorder will be filled. In reality, retailers typically order once prior to the selling season in order to have the product available when customers want it, and then place reorders during the season if early season sales are strong. As discussed in Fisher and Raman, 1996, the information provided by these early season sales dramatically increases the accuracy of the demand forecast. To simplify the presentation of our analysis, we assume that the request for restocking occurs at the end of the selling season, when the realization of demand has been fully observed. Although this eliminates the possibility that a retailer can both receive a second shipment and have excess stock at the end of the season, our model provides insight into the trade-off that the retailer faces between improved demand information versus higher costs and lower certainty of getting what he has ordered.

To analyze the effect of the manufacturer's policy of filling reorders on supply chain performance, we assume that the manufacturer acts as a leader by *announcing* the fraction (α) of reorder requests that she will fill. In practice, such an announcement could be made by establishing a reputation based on long term performance. We further assume that individual requests for restocking are either filled completely or not at all, such that from the perspective of an individual retailer, he will receive all of the units requested in a reorder with probability α, and none of the units with probability $1-\alpha$. This assumption can be justified in terms of two practical considerations. First, if each retailer were allocated some fraction of the amount that he ordered, then there would be an incentive for retailers to inflate their orders. Second, this approach may reduce shipping costs relative to those associated with sending partially filled orders to all retailers.

In response to the manufacturer's order refilling policy (α) and the wholesale price w_1, each retailer i places an initial order, denoted by Q_i. The manufacturer then produces these quantities, at a cost of c_1 per unit, and delivers to the retailers.

After receiving his initial order quantity, each retailer experiences a single period of demand, earning revenue of p per unit sold. If the realization x of demand at retailer i exceeds Q_i, we assume that the excess demand $(x - Q_i)^+$ can be backlogged at a cost of b per unit, and the retailer places a reorder with the manufacturer for the number of units in the backlog.

The manufacturer then produces a fraction, α, of the total amount backlogged by all of the retailers. We assume that the manufacturer fills the fraction α of all reorders in a manner that is perceived as random by the retailers. Note that this could result from either the manufacturer filling all requests for reorders on a randomly chosen set of the products that it produces, or by filling a portion of requests on all products. In other words, we assume that it is not necessary for the manufacturer to fill the fraction α of requests for each realization of demand, so long as she fills the fraction α of requests in expectation. This second production run incurs a cost of c_2 per unit and is sold to each of the retailers at w_2 per unit. The decision variables in the manufacturer's optimization problem are the fraction α of reorders filled and the wholesale price w_2 for reorders.

If a retailer's restocking request is not filled, then he experiences lost sales for the backlogged units. Alternatively, if a retailer receives his requested units, then he earns revenue of p less a backlog cost of b per unit. The last quantity captures the costs associated with special shipping to the customer or services necessary for special delivery. Thus, the backlog cost is not incurred if a retailer's order is not filled by the manufacturer. Other than the backlog cost b there is no other penalty incurred by the retailer for shortages, such as loss of goodwill cost, etc.

Each retailer faces independent identically distributed (i.i.d.) demand that has density $f(x)$, and cumulative distribution function $F(x)$. For simplicity, we assume that the manufacturer has the same information about the distribution of demand as do the retailers.

In order to analyze this model, let us first consider the problem faced by retailer i in determining the appropriate amount to order at the first opportunity. Taking w_1, p and b as given, we can express the expected profits of retailer i as follows:

$$\begin{aligned} R(Q_i, \alpha, w_2) &= -w_1 Q_i + p \int_0^{Q_i} x f(x) dx \\ &+ \alpha(p - w_2 - b) \int_{Q_i}^{\infty} (x - Q_i) f(x) dx + p Q_i \overline{F}(Q_i). \end{aligned} \tag{5.1}$$

where $\overline{F}(Q)$ is the converse cumulative distribution evaluated at Q.

Assuming of course that $0 \le \alpha \le 1$, and $c_2 \le w_2 \le p - b$, it is easy to confirm that (5.1) is concave with respect to Q_i, and that the optimal order quantity for retailer i can be expressed as:

$$Q_i^*(\alpha, w_2) = Q^*(\alpha, w_2) = \overline{F}^{-1} \left(\frac{w_1}{p - \alpha(p - w_2 - b)} \right). \tag{5.2}$$

Observe that the retailers' order quantities are decreasing in α. Thus, as the manufacturer becomes more reliable in responding to restocking requests, the retailers decrease the amount that they order initially and become more apt to require restocking.

Let us now consider the perspective of the manufacturer whose expected profits can be expressed in terms of the retailers' optimal responses to her announced restocking policy:

$$\begin{aligned} M(\alpha, w_2) = N \big[&(w_1 - c_1) Q^*(\alpha, w_2) \\ &+ \alpha(w_2 - c_2) \int_{Q^*(\alpha, w_2)}^{\infty} (x - Q^*(\alpha, w_2)) f(x) dx \big]. \end{aligned} \tag{5.3}$$

Since the manufacturer can induce each retailer to order quantity $Q = Q^*(\alpha, w_2)$ by setting α and/or w_2 appropriately, we can alternatively express her profits as the following function of Q

$$M(Q) = N[(w_1 - c_1)Q + \alpha(w_2 - c_2)\overline{H}(Q)], \tag{5.4}$$

where

$$\overline{H}(Q) = E[(x - Q)^+]$$

is the expected amount of backlogged demand at a given retailer. Note that, since $(x-Q)^+$ is a nonnegative random variable, the expected value can be expressed as $E[(x-Q)^+] = \int_0^\infty P[(x-Q)^+ \geq y]dy$. (Justification is provided by Ross, 1998, among others.) Therefore,

$$\overline{H}(Q) = \int_0^\infty \overline{F}(Q + y)dy = \int_Q^\infty \overline{F}(y)dy. \tag{5.5}$$

2.1 Manufacturer controls only the Reorder Fulfillment Rate

Let us first assume that the manufacturer can control only the rate of fulfilling requests for reorders (α). From (5.2), it can be shown that in order to induce an order quantity of Q, the manufacturer must fill reorders at rate:

$$\alpha(Q) = \frac{1}{p - w_2 - b}\left(p - \frac{w_1}{\overline{F}(Q)}\right). \tag{5.6}$$

Substituting (5.6) into (5.4) and rearranging, we obtain a new expression for the manufacturer's expected profit as a function of the induced order quantity:

$$M(Q) = N\left[(w_1 - c_1)Q + \frac{w_2 - c_2}{p - w_2 - b}\left(p\overline{H}(Q) - w_1\frac{\overline{H}(Q)}{\overline{F}(Q)}\right)\right]. \tag{5.7}$$

This expression allows us to make interesting interpretations of the individual terms. Recall that $\overline{H}(Q)$ is equal to the expected backlog at a given retailer. The term $\overline{H}(Q)/\overline{F}(Q)$ can be interpreted as the conditional expectation of the amount reordered by a retailer, given that his demand exceeds his initial order quantity. Unfortunately, $M(Q)$ is in general neither concave nor convex, as indicated by the following Lemma.

Lemma 5.1 *a) $M(Q)$ is convex (concave) if and only if $p\overline{H}(Q) - w_1\frac{\overline{H}(Q)}{\overline{F}(Q)}$ is convex (concave).*

b) If $\overline{H}(Q)/\overline{F}(Q)$ is concave, then M(Q) is convex.

Proof. Part (a) is immediate. For (b), taking the second derivative of (5.5) with respect to *Q*, we have $\overline{H}''(Q) = f(Q) > 0$. Thus, $\overline{H}(Q)$ is convex, and a sufficient condition for *M*(*Q*) to be convex is for $\overline{H}(Q)/\overline{F}(Q)$ to be concave. ◊

Theorem 5.2 *a) If demand is exponentially distributed so that $f(x) = \lambda e^{-\lambda x}$, then M(Q) is convex. For any pair of wholesale prices (w_1, w_2), the manufacturer will optimally offer total fulfillment $(\alpha = 1)$ of restocking requests if:*

$$\frac{w_2 - c_2}{w_1 - c_1} \geq \frac{(w_2 + b)}{w_1} \ln(\frac{p}{w_2 + b}) \tag{5.8}$$

Otherwise, the manufacturer will optimally offer no fulfillment $(\alpha = 0)$ of restocking requests.
b) If demand is uniformly distributed so that $f(x) = (U - L)^{-1}$ for $x \in [L, U]$, then M(Q) is convex. For any pair of wholesale prices (w_1, w_2), the manufacturer will optimally offer total fulfillment $(\alpha = 1)$ of restocking requests if:

$$\frac{w_2 - c_2}{w_1 - c_1} \geq \frac{2[p - (w_2 + b)](w_2 + b)}{pw_1} \tag{5.9}$$

Otherwise, the manufacturer will optimally offer no fulfillment $(\alpha = 0)$ of restocking requests.

Proof. a) If demand is exponentially distributed with parameter λ, then we have $\overline{F}(Q) = e^{-\lambda Q}$ and $\overline{H}(Q) = e^{-\lambda Q}/\lambda$. Thus, $\overline{H}(Q)/\overline{F}(Q) = 1/\lambda$ and from Lemma 5.1(b) it follows that *M*(*Q*) is convex in Q. Therefore, *M*(*Q*) is maximized by inducing retailers to order at one of the two extreme points, i.e. either:

$$Q^*(0, w_2) = \overline{F}^{-1}\left(\frac{w_1}{p}\right) \quad \text{or} \quad Q^*(1, w_2) = \overline{F}^{-1}\left(\frac{w_1}{w_2 + b}\right).$$

For the exponential distribution with parameter λ, (5.2) can be expressed as:

$$Q^*(\alpha, w_2) = -\frac{1}{\lambda} \ln\left(\frac{w_1}{p - \alpha(p - w_2 - b)}\right). \tag{5.10}$$

We can now substitute for Q and $\overline{H}(Q)$ in (5.7) to obtain:

$$M(Q^*(\alpha, w_2)) = N\left[-(w_1 - c_1)\frac{1}{\lambda}\ln\left(\frac{w_1}{p - \alpha(p - w_2 - b)}\right) + \frac{\alpha(w_2 - c_2)w_1}{\lambda(p - \alpha(p - w_2 - b))}\right]. \quad (5.11)$$

From this expression, we can substitute $\alpha = 1$ and $\alpha = 0$, and rearrange the terms to see that:

$$M(Q^*(1, w_2)) - M(Q^*(0, w_2)) = -\frac{N}{\lambda}\left[(w_1 - c_1)\ln\left(\frac{p}{w_2 + b}\right) - \frac{w_1(w_2 - c_2)}{w_2 + b}\right]. \quad (5.12)$$

It is easy to see that the above expression is non-negative if and only if (5.8) is true.
b) If demand is uniformly distributed on (L,U), then we have:

$$\overline{F}(Q) = \frac{U - Q}{U - L} \text{and } \overline{H}(Q) = \frac{(U - Q)^2}{2(U - L)}.$$

Thus, $\overline{H}(Q)/\overline{F}(Q) = (U - Q)/2$ and from Lemma 5.1(b) it follows that $M(Q)$ is convex in Q. Therefore, $M(Q)$ is maximized by inducing retailers to order at one of the two extreme points, i.e. either:

$$Q^*(0, w_2) = U - \frac{w_1}{p}(U - L) \text{ or } Q^*(1, w_2) = U - \frac{w_1}{w_2 + b}(U - L).$$

Substituting into (5.7) and rearranging, we have:

$$M(Q^*(\alpha, w_2)) = N\left[(w_1 - c_1)(U - \frac{w_1(U - L)}{p - \alpha(p - w_2 - b)}) + (w_2 - c_2)\frac{\alpha w_1^2(U - L)}{2(p - \alpha(p - w_2 - b))^2}\right]. \quad (5.13)$$

From this expression, we can substitute $\alpha = 1$ and $\alpha = 0$ to see that:

$$M(Q^*(1, w_2)) - M(Q^*(0, w_2)) = N\left[(w_1 - c_1)\frac{w_2 + b - p}{p(w_2 + b)} + \frac{(w_2 - c_2)w_1}{2(w_2 + b)^2}\right](U - L)w_1. \quad (5.14)$$

By rearranging the terms inside the bracket, it is easy to see that the above expression is non-negative if and only if (5.9) is true. ♢

It is of interest that for both uniformly and exponentially distributed demand, the manufacturer's optimal policy fills all requests for restocking if and only if the ratio of profit margins between the restocking and the initial ordering opportunity exceeds a certain threshold.

Although the above analytical results do not extend to the gamma or normal distributions, which are often used to model demand, we have performed an extensive set of numerical experiments for these distributions. These experiments indicated that for the gamma distribution $M(Q)$ is not always convex, but nevertheless it is maximized by either filling all requests for restocking or by filling none of them. This also appears to be true when demand is normally distributed. Although we were able to construct some examples with normally distributed demand in which the profit was maximized for some value $0 < \alpha < 1$, in all such cases the manufacturer's profit was practically constant in α, with a total variation of less than 1% between the minimum and maximum values. Thus, although the specifics of our analytical results do not extend to the gamma or normal distributions, the basic conclusions carry through.

2.2 Offering Re-Stocking as a Benefit to Retailers

In many situations, a manufacturer who is introducing a restocking option must guarantee that a retailer can do at least as well after the introduction of restocking, regardless of whether he chooses to take advantage of the option of restocking. For example, there may be many retailers who are either unable or are unwilling to change their own business practices in order to take advantage of restocking. In order to avoid disenfranchising these retailers when a restocking policy is introduced, the manufacturer may want to avoid changing the wholesale price (w_1) that she charges for the initial orders that are placed before observing demand. Moreover, by leaving this early wholesale price unchanged after introducing restocking, the manufacturer signals retailers that the new policy can only benefit them.

Lemma 5.3 *Taking w_1 as fixed, and assuming that $p > w_1 \geq c_1$ and $p - b \geq c_2 \geq c_1$, there is always a total reorder fulfillment policy $(\alpha = 1)$ for which both the manufacturer and the retailers' expected profits are at least as large as without reorder fulfillment.*

To see that this is true, note that by setting $\alpha = 1$ and $w_2 = p - b$, the retailers are indifferent between filling backorders through restocking and experiencing lost sales. Thus, this policy induces the retailers to make the same initial orders as if there were no reorder fulfillment. The retailers earn the identical profit as they would without the restocking option, and any reorder requests that the manufacturer receives increase her profits.

The above Lemma also implies that, if the manufacturer has control over the wholesale price for restocking, she will always prefer $\alpha = 1$ to $\alpha = 0$. As long as *M*(*Q*) is convex, as we have shown it to be for uniform and exponential demand distributions, the manufacturer's optimal policy will be to set $\alpha = 1$ and find the profit maximizing restocking wholesale price $w_2^*(\alpha = 1)$. On the other hand, both for uniform and exponential demand distributions $M(Q^*(\alpha, w_2))$ is unimodal in w_2. This is so because it can be shown that if the partial derivative $\partial M(Q^*(\alpha, w_2))/\partial w_2 < 0$ for some w_2, then it remains negative for all $w_2' \geq w_2$. This gives rise to the following Theorem, the proof of which we have omitted:

Theorem 5.4 *a) If demand is exponentially distributed with parameter λ, then the manufacturer's optimal restocking policy is to set $\alpha = 1$ and*

$$w_2^* = \begin{cases} p - b & \text{if } 2w_1 - c_1 - \frac{\alpha w_1(p-c_2-b)}{\lambda p} \geq 0 \\ \frac{\lambda(2w_1-c_1)[p(1-\alpha)+\alpha b]+\alpha w_1 c_2}{\alpha(w_1-\lambda(2w_1-c_1))}, & \text{otherwise} \end{cases}.$$

b) If demand is uniformly distributed on (L, U), then the manufacturer's optimal restocking policy is to set $\alpha = 1$ and:

$$w_2^* = Min\left\{p - b, \frac{(3w_1 - 2c_1)b + 2w_1c_2}{2c_1 - w_1}\right\}.$$

2.3 The Role of Re-stocking in Channel Coordination

In this section, we address the issue of the extent to which the introduction of a restocking policy can or will serve to coordinate the supply chain. If the manufacturer and retailer interact only once, prior to the observation of demand, the quantity that maximizes the channel profits satisfies:

$$\overline{F}(Q) = \frac{c_1}{p}. \tag{5.15}$$

This is the solution to a standard newsvendor problem with the marginal cost of production (c_1) as the cost of overage, and the profit margin $(p - c_1)$ as the cost of underage. The quantity ordered by the retailer will satisfy:

$$\overline{F}(Q) = \frac{w_1}{p}. \tag{5.16}$$

Clearly, the only way that the manufacturer can induce the retailer to order the channel coordinating quantity is by setting w_1 so that (5.15) is equal to (5.16), which implies that the channel can be coordinated only when $w_1 = c_1$. Hence, in this environment, we cannot both coordinate the channel and allow the manufacturer to earn positive profit.

However, as shown in the following theorem, this is not the case if the manufacturer introduces a restocking policy. Indeed, with a restocking policy, the manufacturer can coordinate the channel with an early wholesale price (w_1) that is strictly greater than her marginal costs.

Theorem 5.5 *As long as $p - b \geq c_2$, the manufacturer can coordinate the channel, i.e. maximize channel profits, by setting $\alpha = 1$ and setting w_1 and w_2 such that:*

$$w_1(Q_c) = c_1 \frac{w_2 + b}{c_2 + b}, \tag{5.17}$$

where Q_c is the order quantity that maximizes channel profits.

Proof. From the channel perspective it can never be profitable to let backorders go unfilled so long as $p - b \geq c_2$, which implies that $\alpha = 1$ is optimal. Therefore, the initial order quantity is the solution to a newsvendor problem in which the cost of over-production is c_1, and the cost of under-production is $c_2 - c_1 + b$. Thus, the channel coordinating order quantity must satisfy:

$$(Q_c) = \overline{F}^{-1}\left(\frac{c_1}{c_2 + b}\right). \tag{5.18}$$

The relationship in (5.17) follows from setting the retailer's order quantity as defined in (5.2) equal to the channel coordinating quantity in (5.18). ◇

This result is significant because it implies that the introduction of a restocking policy allows the manufacturer to use a linear pricing policy to both coordinate the channel and extract a share of the profits from the retailer, which is not true in the absence of a restocking policy.

3. Numerical Analysis

3.1 Total Supply Chain Profits

The analytical results derived in Section 2 characterize how the manufacturer's profits depend on α, the portion of reorders filled, and w_2, the wholesale price per unit for reorders. In this section, we perform a numerical study to better understand the effects of partial replenishment policies on the performance of the channel. Even in a simple case

	Varying α			Varying w_2
Case	1	2	3	4
Distribution	U(100,1000)	U(100,1000)	Exp(1/100)	U(100,500)
p	200	100	50	50
b	0	3	1	7
c_1	15	20	6	20
c_2	40	30	29	26
w_1	49	35	25	22
	w_2=50	w_2=50	w_2=37	α=1

Table 5.1. Data for numerical examples

where the manufacturer's profits are increasing in w_2 and the retailer's profits are decreasing in w_2, the effect on channel profits is ambiguous. Wholesale prices have only second order effects on channel profits via their effects upon the initial order of the retailer. Changes in the portion of reorders filled has both first and second order effects on channel profits.

To demonstrate the relationship between restocking policies that maximize the manufacturer's profits versus those that maximize channel profits, we provide the following numerical examples. In these examples we have considered four cases for the demand distribution and the various parameter values. The data used for each case are summarized in Table 1.

Figure 5.1 shows the manufacturer and channel profits as a function of α for Case 1. In this case, the manufacturer's profits are strictly decreasing in α. Although the channel profits decrease and then increase in α, both the profits of the manufacturer and of the channel are maximized at $\alpha = 0$.

Figure 5.2 shows the manufacturer and channel profits as a function of α for Case 2. In this case the manufacturer will choose not to offer a reorder opportunity. However, we can see in Figure 5.2 (b) that channel profits are increasing in α throughout the range. Therefore, channel profits would be higher if a reorder opportunity was offered.

An example with exponential distribution also shows behavior in which the manufacturer's profits are decreasing then increasing, while the channel profits increase throughout the range, as is shown in Figure 5.3 for Case 3. Note, however, that in this case the manufacturer will still choose the channel optimal policy of $\alpha = 1$.

We also consider the effects of selecting w_2 for a given α value. Figure 5.4 shows the manufacturer and channel profits as a function of w_2 for Case 4. We find again that the channel profits are strictly increasing or decreasing in many of the cases. However, there are sometimes discrep-

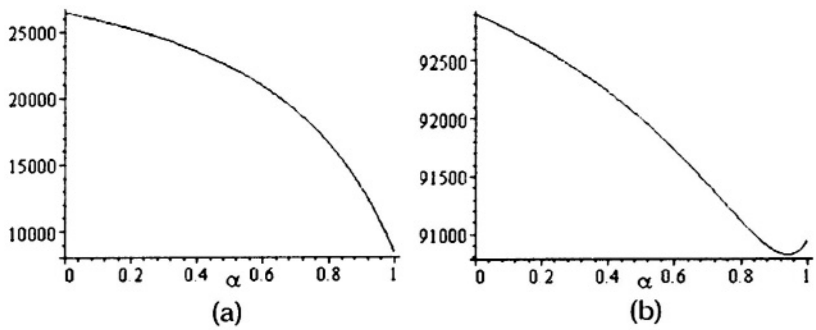


Figure 5.1. Manufacturer (a) and Channel Profit (b) - Case 1

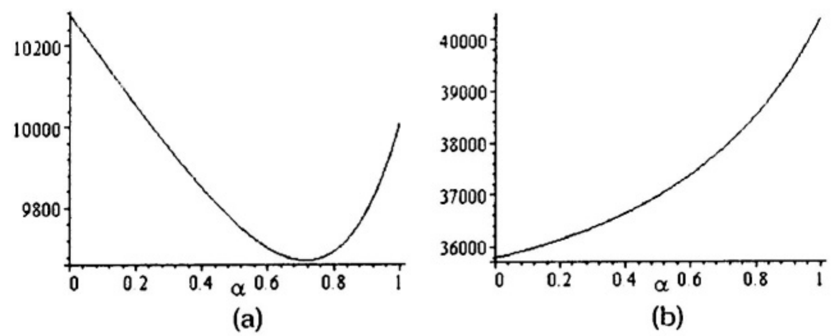


Figure 5.2. Manufacturer (a) and Channel Profit (b) - Case 2

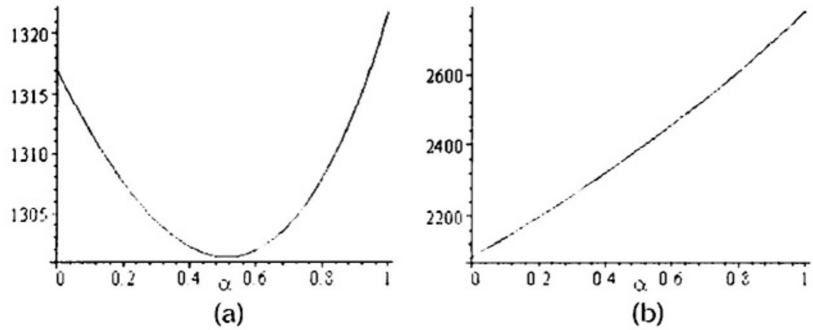


Figure 5.3. Manufacturer (a) and Channel Profit (b) - Case 3

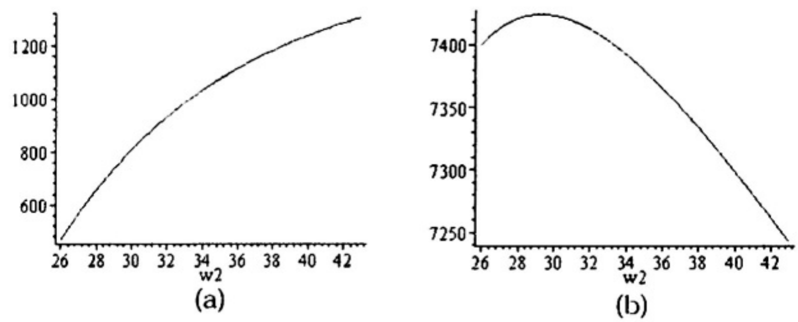


Figure 5.4. Manufacturer (a) and Channel Profit (b) - Case 4

ancies between what is best for the manufacturer and what is best for the channel. Indeed for the case shown in Figure 5.4, the total channel profits are maximized when $w_2 \approx 29$, whereas the manufacturer's profits are increasing over the entire range. As a result, the manufacturer will tend to charge a larger wholesale price than would be optimal from the perspective of the channel. Note that this is consistent with the results of Jeuland and Shugan, 1983, who showed that in a bi-lateral monopoly, both the manufacturer and the retailer will tend to set higher margins than would be required to maximize channel profits.

3.2 Single Production - Partial Fulfillment Model

In many practical situations, manufacturers lack the operational flexibility to respond to requests for restocking with additional production. In such cases, if the product being sold is standardized, manufacturers often produce more than enough to satisfy retailers' initial orders in anticipation of requests for restocking. In this section we perform a numerical investigation to better understand the effects of such overproduction.

As before, we assume that there are N retailers facing independent, identically distributed demand, and that the manufacturer fills requests for restocking without regard for the number of units ordered, and that each request is either filled completely or not at all. We assume that N is large enough that the central limit theorem can be applied and that the distribution of the combined requests for restocking from the retailers can be represented by a normal distribution.

In contrast to our original model, we now assume that the manufacturer produces only once, prior to any observations of end demand. In particular, we assume that, after receiving the retailers' orders, the manufacturer produces enough to fill these orders and to fill some portion of the anticipated demand for restocking. Let $y_i(Q)$ denote the number of units of demand backlogged at retailer i, given that he ordered Q units initially. Then $y_i(Q), i = 1, \ldots, N$ are independent identically distributed random variables with mean and variance

$$\begin{aligned} E[y_1(Q)] &= \int_Q^\infty (x - Q) f(x) dx, \\ \sigma^2_{y_1(Q)} &= \int_Q^\infty (x - Q)^2 f(x) dx - E[y_1(Q)]^2. \end{aligned}$$

Let $y_T(Q)$ denote the total quantity requested for restocking by all retailers given that they have each ordered Q units initially. By the central limit theorem, $y_T(Q)$ is approximately normally distributed with the following mean and variance:

$$\begin{aligned} E[y_T(Q)] &= NE[y_1(Q)], \\ \sigma^2_{y_T(Q)} &= N\sigma^2_{y_1(Q)}. \end{aligned}$$

Thus, in order to insure that the expected amount of unsatisfied demand for restocking is equal to $1 - \gamma$ of the total expected demand for restocking, the manufacturer must produce:

$$T = NQ + E[y_T(Q)] + z_\gamma \sigma_{y_T(Q)},$$

where

$$z_\gamma = L^{-1}\left(\frac{(1-\gamma)E[y_T(Q)]}{\sigma_{y_T(Q)}}\right),$$
$$L(t) = \int_t^\infty (x-t)\phi(x)dx,$$

and $\phi(x)$ is the density of the standard (unit) normal distribution. Assuming that any shortages are allocated randomly among the retailers and that they cannot misrepresent their backlogs, then γ is a good approximation to α. In other words, by filling a fraction γ of the demand for restocking, the manufacturer can, on average, fill the fraction α of the requests.

From the retailer's perspective, the manufacturer's policy results in requests for restocking being filled with probability $\alpha = \gamma$, and each retailer orders the quantity identified in $Q^*(\gamma, w_2)$, where γ replaces α:

$$Q^*(\gamma, w_2) = \overline{F}^{-1}\left(\frac{w_1}{p-\gamma(p-w_2-b)}\right). \tag{5.19}$$

Note, however, that when the manufacturer's initial production quantity is in excess of the combined initial orders from the retailers, the manufacturer exposes herself to some risk, and may incur the cost of unsold product. As a result, it will become prohibitively expensive for her to guarantee complete fullfillment of reorders, i.e. $\gamma = 1$. Specifically, the cost to achieve very high levels of service becomes prohibitive, due to the large "safety stock", represented by the term $z_\gamma\sigma_{y_T(Q)}$, that is required in this case. The manufacturer's profits are shown in Figure 5.5 for γ from 0 to 1, for a case where each retailer's demand follows uniform distribution between 100 and 1000, and $N = 100, p = 100, b = 3, c = 20, w_1 = 35, w_2 = 50$. This case is the same as Case 2 in Table 1, with the only difference being that all production takes place in the first period at cost c_1. There are two things worth noting: First, it can be seen that, as expected, the profits show an abrupt decrease for values of γ close to 1 due to the prohibitive cost of producing enough in advance to guarantee fulfillment of all reorder requests. Second, the manufacturer's profits are maximized at a value of γ that is very close to 1. Recall that when reorders had to be filled from higher cost production the manufacturer's profits were maximized by filling no reorders, as shown in Figure 5.2. The main reason for this difference is that when products are standardized and the manufacturer can use the inexpensive early production to build inventory, she can earn a larger margin on the reorders. Thus, she may encourage reorders when she can fill them from

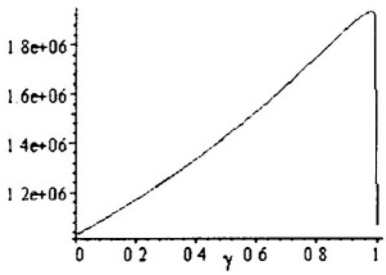


Figure 5.5. Manufacturer's Profits in Single Production Partial-Replenishment

inventory, even though she would discourage them if she had to employ an expensive mode of production to satisfy them.

4. Conclusions and Extensions

We have developed models for partial quick response policies used in practice, and identified situations in which such policies are beneficial not only to retailers but also to manufacturers.

For the uniform and exponential distributions, we analytically characterized the optimal policies when the portion of reorders and/or the wholesale price per unit of reorders are under the manufacturer's control. For a given wholesale price, it is optimal to either fill all reorders or not offer a reorder opportunity. It is optimal to offer a reorder opportunity when the ratio of the manufacturer's profit margin on reorders to her margin on the initial orders exceeds a certain threshold. For a given portion of reorders filled, there is an optimal reorder wholesale price that exists in the interval between the manufacturer's production cost at the reorder opportunity, and the value at which the retailer would be indifferent to participating. Combining these results, we showed that if the manufacturer controls both the portion of reorders filled and the wholesale price charged for reorders, the optimal policy is to fill all reorders and charge the appropriate maximizing wholesale price for complete fulfillment.

We numerically explored this policy's effect on the channel profits, and found that the policy that maximizes the manufacturer's profits does not necessarily coincide with the one that maximizes channel profit.

Therefore, without some additional mechanism, the manufacturer would not always have an incentive to coordinate the channel.

In our numerical experiments, we also considered a related model of partial replenishment in which the manufacturer lacks access to reactive production but can build inventory before the selling season. Relative to the case where she cannot build inventory, but can produce reactively, this allows the manufacturer to utilize low cost production to satisfy reorders. On the other hand, it can result in the manufacturer's producing units that retailers do not need. There are a number of questions that would be worth pursuing in this line of inquiry. For example, it would be interesting to know how the manufacturer's optimal production quantity in anticipation of reorders would compare to the optimal quantity for a vertically integrated channel. It would also be interesting to investigate how a manufacturer would use a combination of inventory and responsive production to satisfy retailers' reorders.

Other directions in which this work could be extended include different cost structures that take into account set-up and holding costs within the period. Additionally, it could be useful to analyze the following trade-off: as a retailer postpones making a request for a reorder, he gains more information about demand, but incurs a greater risk that the manufacturer will be unable to fulfill his request. It might also be of interest to study the effects of competition, at either the retailer or manufacturer levels. Finally, it will also be important to consider the managerial issues of cooperation necessary to implement the policies described here.

References

Eppen, G. D. and Iyer, A. V. (1997). Backup agreements in fashion buying - the value of upstream flexibility. *Mgt. Sci.,* 43(11):1469–1484.

Fisher, M. and Raman, A. (1996). Reducing the cost of demand uncertainty through accurate response to early sales. *Oper. Res.,* 44(1):87–99.

Hammond, J. H. and Raman, A. (1996). Sport obermeyer, ltd. Harvard Business Case No. 695022.

Iyer, A. V. and Bergen, M. E. (1997). Quick response in manufacturer-retailer channels. *Mgt. Sci.,* 43(4):559–570.

Jeuland, A. P. and Shugan, S. (1983). Managing channel profits. *Mktg. Sci.,* 2(3):239–272.

Lau, H. and Lau, A. H. (1997). A semi-analytical solution for a newsboy problem with mid-period replenishment. *J. Opl. Res. Soc.,* 48(12): 1245–1253.

Ross, S. (1998). *A First Course in Probability* 5th ed., Prentice Hall, Upper Saddle River, NJ.

Signorelli, S.and Heskett, J. L. (1984). Benetton. Harvard Business Case No. 685020.

Chapter 6

USING REVENUE SHARING TO ACHIEVE CHANNEL COORDINATION FOR A NEWSBOY TYPE INVENTORY MODEL

Barry Alan Pasternack
Department of Information Systems and Decision Sciences
CSU Fullerton
Fullerton, CA 92834
bpasternack@fullerton.edu

Abstract This chapter considers a single period inventory (newsboy) problem in which a manufacturer will both sell the item to the vendor outright as well as offer the item to the vendor on a revenue sharing (consignment) basis. In the latter case, the amount of money the vendor pays per unit is less than if the item is purchased by the vendor, but the vendor must share some of the revenue with the manufacturer. The purpose of our analysis is to investigate the effect of such a strategy on the vendor's purchasing decision and demonstrate how such a revenue sharing scheme can be used to achieve channel coordination.

1. Introduction

This chapter investigates the situation in which a vendor, in a newsboy-type situation, has the option of purchasing the item outright and/or obtaining the item through a revenue sharing agreement with the manufacturer. Of specific interest is whether revenue sharing can be used to achieve channel coordination.

The motivation for this work comes from an article in the March 25, 1998 edition of the *Wall Street Journal* dealing with the concept of revenue sharing at video rental stores. The article described efforts by Blockbuster Video to obtain tapes from Hollywood studios on a revenue sharing basis. The article cited research by Time Warner that indicated 20% of customers were unable to obtain their first choice video due to a

J. Geunes et al. (eds.),
Supply Chain Management: Models, Applications, and Research Directions, 117–136.

stock out situation. To remedy this situation, stores were being offered videotapes for $8 each on a revenue sharing basis versus $65 each for outright purchase.

The purpose of this chapter is to examine the strategy of revenue sharing for a newsboy (single period) inventory model. The fact that the problem faced by videotape rental stores is analogous to the newsboy model was shown in Drezner and Pasternack [1999]. We note, however, that while our motivation for studying this problem originated with videotape rental stores, the results of this work are applicable to any newsboy inventory situation and may, for example, be of interest to publishers of periodicals or manufacturers of perishables such as baked goods.

The newsboy (single period) inventory model is covered in most introductory management science texts (see for example Lawrence and Pasternack 1998). There have also been numerous papers in the literature dealing with extensions to this model. Silver, Pyke, and Peterson [1998] gave a partial review of the literature in their textbook. Khouja [1999] presented an extensive review of the literature dealing with such extensions. In this paper Khouja reviewed some 90 publications and classified them into eleven categories based on the type of extension to the classical newsboy problem.

Following Khouja's classification, the material in this chapter would fall into the category of problems dealing with different supplier pricing policies. Extensions to the newsboy problem in terms of different supplier pricing policies have focused on quantity discounts (see for example Jucker and Rosenblatt 1985, and Lin and Kroll 1997), permitting emergency supplies obtainable at a premium cost over the original cost per unit (see for example, Khouja 1996), and multiple suppliers (see for example, Kabak and Weinberg 1972).

As with the literature on the single period inventory problem, there are a number of papers that have dealt with pricing policies to achieve channel coordination (an excellent review of this literature can be found in Tsay, Nahmias, and Agrawal 1999). For example, Jeuland and Shugan [1983] and Monahan [1984] both considered discounting schemes that achieved channel coordination. Lal [1990] looked at franchise fees and monitoring schemes and their effect on channel coordination. These papers however, did not deal with a newsboy type situation. For a newsboy type good, Pasternack [1985] showed that a manufacturer could set a pricing and return policy so that channel coordination could be achieved. Lau and Lau [1999] extended Pasternack's work to account for the manufacturer's and retailer's attitude towards risk with a goal of maximizing the manufacturer's profit. Padmanabhan and Png [1997] developed a single period inventory model in which demand is linearly related to price and examined the effect that full returns and no returns have on retailer strategies and the manufacturer's profit. Lariviere and Porteus [1999] considered a newsboy type inventory model and looked at structural issues for the manufacturer's pricing policy as well as the effect that market size and variability have on the pricing decision. Brown and Lee [1998] examined a situation faced in the semiconductor industry in which a vendor could both purchase units outright from a manufacturer as well as purchase options for additional units. These options could be exercised at a future point in time after additional information for the good's demand has been observed. They considered situations in which several of the costing parameters could change and showed that, for their model, channel coordination can be attained to allow greater profit for both the manufacturer and the supplier. While the model considered by Brown and Lee has similarities to the one presented in this chapter, their model did not include the possibility of the vendor incurring a goodwill cost due to unavailability of the item at the retail level. As we will show, the inclusion of goodwill cost may lead to a situation in which it is impossible for the manufacturer to gain in a channel coordination scheme without the use of a side payment from the vendor to the manufacturer.

The focus of this chapter is on the vendor's optimal procurement strategy and on the manufacturer's pricing strategy for offering items on consignment in order to achieve channel coordination. We allow the vendor to obtain the item through two different costing schemes. In particular, we assume that the manufacturer has been selling the item to the vendor on a conventional basis and now wishes to also offer the good to the vendor on a revenue sharing (consignment) basis. Hence, the vendor will have the choice of purchasing the item outright, obtaining the item on consignment, or doing a combination of both. As with much of the literature on channel coordination, we will assume that the manufacturer behaves like a Stackelberg leader and that the vendor acts to maximize its expected profits. Our analysis will examine the vendor's procurement strategy and how a consignment scheme can achieve channel coordination.

The chapter is divided into four sections. In Section 2 we introduce the general model. We show conditions under which it is optimal for the vendor to only purchase the item outright, only obtain the item on consignment, or obtain the item through both outright purchase and consignment. In Section 3 we discuss the issue of channel coordination. We show that a consignment scheme can be used to achieve channel coordination and present pricing formulas that a manufacturer can use to “force” the vendor to order, in total, what is in the best interests of the channel. While this may likely be the most compelling reason for manufacturers to adopt this strategy, we show that it is possible for the manufacturer’s expected profit to decline as a result of such coordination. Formulas for the vendor’s and manufacturer’s expected profit are presented and examples are given for the case of a uniform demand distribution. Section 4 contains some concluding remarks and potential avenues for future research.

2. Optimal Strategies When the Vendor’s Available Funds Are Unlimited

In this section we introduce the general model and investigate the vendor’s optimal strategy. We assume that there are no limits on the funds the vendor has available for procurement and that either the cost of capital is low or the time horizon under consideration is short. (For an analysis of the vendor’s optimal purchase/consignment strategy when funding is limited see Pasternack [2000].) We note that because a vendor does not have to obtain the item on consignment, in order for any consignment scheme to be attractive to a vendor, it must result in a higher expected profit than if the vendor obtains all units through outright purchase.

The notation we use is as follows:

- c_1 = the vendor’s cost per unit if the vendor obtains the item from the manufacturer through outright purchase.
- p_1 = the retail price per unit and therefore the vendor’s revenue per unit if the vendor obtains the item from the manufacturer though outright purchase.
- s = the vendor’s salvage value per unit if the vendor obtains the item from the manufacturer through outright purchase.
- g = the vendor’s goodwill cost per unit if the vendor is out of stock of the item.
- m = the manufacturer’s production cost per unit.
- c_2 = the vendor’s cost per unit if the vendor obtains the item from the manufacturer on consignment.
- p_2 = the vendor’s revenue per unit if the vendor obtains the item from the manufacturer on consignment. (Note that $p_1 - p_2$ equals the revenue paid to the manufacturer if an item purchased on consignment is sold by the vendor.)

- Q_1 = the number of units the vendor obtains from the manufacturer through outright purchase.
- Q_2 = the number of units the vendor obtains from the manufacturer on consignment.
- Q = the number of units ordered if the supply channel is coordinated.
- $EV(Q_1, Q_2)$ = the vendor's expected profit if it purchases Q_1 units and obtains Q_2 units on consignment.
- $EM(Q_1, Q_2)$ = the manufacturer's expected profit if the vendor purchases Q_1 units and obtains Q_2 units on consignment.
- $EP(Q)$ = the total expected channel profit if the channel is coordinated and the order quantity is Q.
- Q_1^* = the optimal number of units the vendor should obtain from the manufacturer through outright purchase if the vendor wishes to maximize its expected profit.
- Q_2^* = the optimal number of units the vendor should obtain from the manufacturer on consignment if the vendor wishes to maximize its expected profit.
- Q^* = the optimal number of units ordered for a coordinated supply channel.
- $f(x)$ = the probability density function of demand.

We will assume that due to competitive pressure, the values of $c_1, p_1, s, g,$ and m are fixed[1], however the manufacturer has control over setting the values of c_2 and p_2. Given the cost structure set by the manufacturer, the vendor will then determine the order quantity that maximizes its expected profit.

Following this notation, we see that if the vendor purchases the item outright from the manufacturer, the vendor earns a gross profit of $p_1 - c_1$ for each unit sold and the manufacturer earns a gross profit of $c_1 - m$ for each unit ordered by the vendor. If however, the vendor orders the item from the manufacturer on a consignment basis, the vendor would

[1]For example, videotapes are not only sold to stores that rent tapes, but also to stores which sell tapes outright to consumers. Even among stores that rent tapes, some may not wish to participate in a consignment scheme. For these reasons it is assumed that the vendor's cost per unit, c_1, and the retail price per unit, p_1, are not affected by the introduction of the consignment scheme.

pay the manufacturer c_2 per unit plus an amount equal to $p_1 - p_2$ for each unit sold. As a result, for each unit purchased on consignment and sold by the vendor, the vendor earns a gross profit of $p_2 - c_2$ and the manufacturer earns a gross profit of $p_1 - p_2 + c_2 - m$.

We further assume that if the vendor purchases the item outright from the manufacturer the salvage value, s, from each unsold unit accrues to the vendor, whereas if the item is obtained from the manufacturer on a consignment basis the salvage value accrues to the manufacturer. In all cases the goodwill cost per unit due to shortage, g, is assumed to accrue to the vendor.

The following conditions are assumed to hold regarding p_1, p_2, c_1, c_2, and s.

- $p_1 > p_2$ (The vendor's revenue per unit is greater if it purchases the item than if it obtains the item on consignment.)
- $c_1 > c_2$ (The vendor's cost per unit is greater if it purchases the item than if obtains the item on consignment.)
- $c_1 > s$ (The vendor's cost per unit for purchasing the item is greater than its salvage value)
- $p_1 > c_1$ (For items purchased, the vendor's revenue per unit is greater than the cost per unit.)
- $p_1 - p_2 > c_1 - c_2$ (For it to be worthwhile for the manufacturer to offer the item to the vendor on a consignment basis, the manufacturer's gross profit per unit from consignment should be greater than the gross profit from outright sale to the vendor. This also states that the vendor's gross profit per unit from outright purchase is greater than if it obtains the item on consignment.)

Note that we will not require $p_2 > c_2$. That is, for items obtained on consignment, the vendor's revenue per unit may be less than the cost per unit. While such a situation would rarely arise, it is conceivably possible that it may be worthwhile for the vendor to obtain goods on consignment even if they would be sold at a loss in order to avoid the potential of incurring extremely high goodwill costs.

Operationally, because a vendor earns a larger gross profit on the units it obtains through outright purchase than on consignment, we assume that the vendor will first sell the units it purchases outright and will only sell those it obtains on consignment after all purchased units have been sold. The basis for our analysis is to investigate the Karush-Kuhn-Tucker (KKT) conditions in order to determine optimal strategies. Based on the above notation, we have:

$$EV(Q_1, Q_2) = \int_0^{Q_1} [p_1 x + s(Q_1 - x)] f(x) dx + \int_{Q_1}^{Q_1+Q_2} [(p_1 - p_2)Q_1 + p_2 x] f(x) dx + \int_{Q_1+Q_2}^{\infty} [p_1 Q_1 + p_2 Q_2 - g(x - Q_1 - Q_2)] f(x) dx - c_1 Q_1 - c_2 Q_2 \quad (6.1)$$

and

$$EM(Q_1, Q_2) = \int_0^{Q_1} s Q_2 f(x) dx + \int_{Q_1}^{Q_1+Q_2} [(p_1 - p_2)(x - Q_1) + s(Q_1 + Q_2 - x)] f(x) dx + \int_{Q_1+Q_2}^{\infty} (p_1 - p_2) Q_2 f(x) dx + (c_1 - m)Q_1 + (c_2 - m)Q_2. \quad (6.2)$$

The first term in equation (6.2) arises due to the fact that any items ordered by the vendor on consignment and unsold are returned to the manufacturer who will then dispose of them for their salvage value.

The problem faced by the vendor is therefore:

$$\textit{Maximize} \;\; EV(Q_1, Q_2) \quad (P1)$$

$$\text{s.t.} \quad -Q_1 \leq 0, \quad -Q_2 \leq 0.$$

The partial derivatives of $EV(Q_1, Q_2)$ are as follows:

$$\frac{\partial EV(Q_1, Q_2)}{\partial Q_1} = F(Q_1)(s - p_1 + p_2) - F(Q_1 + Q_2)(p_2 + g) - c_1 + p_1 + g \quad (6.3)$$

and

$$\frac{\partial EV(Q_1, Q_2)}{\partial Q_2} = [1 - F(Q_1 + Q_2)](p_2 + g) - c_2 \quad (6.4)$$

The following KKT conditions (see Hillier and Lieberman [2001]) are therefore required for optimality:

(1) $Y_1 Q_1^* = 0$

(2) $Y_2 Q_2^* = 0$

(3) $F(Q_1^*)(s - p_1 + p_2) - F(Q_1^* + Q_2^*)(p_2 + g) - c_1 + p_1 + g = -Y_1$

(4) $[1 - F(Q_1^* + Q_2^*)](p_2 + g) - c_2 = -Y_2$

(5) $Y_1 \geq 0, Y_2 \geq 0, Q_1^* \geq 0, Q_2^* \geq 0$

where Y_1 and Y_2 are the Lagrange multipliers (dual variables) for this problem.

We also see that:

$$\frac{\partial^2 EV(Q_1, Q_2)}{\partial Q_1^2} = f(Q_1)(s - p_1 + p_2) - f(Q_1 + Q_2)(p_2 + g) \quad (6.5)$$

$$\frac{\partial^2 EV(Q_1, Q_2)}{\partial Q_2^2} = -f(Q_1 + Q_2)(p_2 + g) \quad (6.6)$$

and

$$\frac{\partial^2 EV(Q_1, Q_2)}{\partial Q_1 \partial Q_2} = -f(Q_1 + Q_2)(p_2 + g). \quad (6.7)$$

By looking at the second-order partial derivatives for $EV(Q_1, Q_2)$ we note that if $p_1 - p_2 - s \geq 0$, then $EV(Q_1, Q_2)$ is concave.

Our interest lies in determining the structure of the optimal strategy for the vendor. This gives rise to the following three theorems.

Theorem 6.1 *It is impossible for* $Q_1^* = 0$ *and* $Q_2^* > 0$

Proof: If $Q_2^* > 0$ then from KKT condition (2) $Y_2 = 0$. Hence, from KKT condition (4) we have:

$$F(Q_1^* + Q_2^*) = (p_2 + g - c_2)/(p_2 + g). \quad (6.8)$$

Substituting equation (6.8) into KKT condition (3) gives:

$$F(Q_1^*)(s - p_1 + p_2) + (p_1 - p_2 - c_1 + c_2) = -Y_2. \quad (6.9)$$

But if $Q_1^* = 0$, then $F(Q_1^*) = 0$ and it would be impossible for equation (6.9) to be satisfied.

Theorem 6.1 states that it would never pay for the vendor to obtain the item from the manufacturer only on a consignment basis.

Theorem 6.2 *If* $\frac{p_2+g}{c_2} > \frac{p_1+g-s}{c_1-s}$*, then it is impossible for* $Q_1^* > 0$ *and* $Q_2^* = 0$.

Proof: Suppose $Q_1^* > 0$. Then from KKT condition (1) we have $Y_1 = 0$. If we also assume that $Q_2^* = 0$, from KKT condition (4) we have:

$$F(Q_1^*) = \frac{p_2+g-c_2+Y_2}{p_2+g} \tag{6.10}$$

while from KKT condition (3) we have:

$$F(Q_1^*) = \frac{p_1+g-c_1}{p_1+g-s}. \tag{6.11}$$

But, if $\frac{p_2+g}{c_2} > \frac{p_1+g-s}{c_1-s}$, then $\frac{p_2+g-c_2+Y_2}{p_2+g} \geq \frac{p_2+g-c_2}{p_2+g} > \frac{p_1+g-c_1}{p_1+g-s}$.

Since we know from Theorem 6.1 that obtaining the good only on consignment cannot be optimal, Theorem 6.2 gives conditions under which it is optimal to both purchase the item outright and obtain it by consignment. In essence, what Theorem 6.2 states is that if the ratio of revenue to cost for goods obtained on consignment is high enough, it will never be optimal for the vendor to only purchase the good. Such a condition will occur if the salvage value is low and the goodwill cost is high[2]. In such cases the optimal amount for the vendor to purchase and to obtain on consignment can be determined by recognizing that $Y_1 = Y_2 = 0$ and solving for KKT conditions (3) and (4). This gives the following relationships for Q_1^* and Q_2^*.

$$F(Q_1^*+Q_2^*) = \frac{p_2+g-c_2}{p_2+g} \tag{6.12}$$

and

$$F(Q_1^*) = \frac{p_1-p_2+c_2-c_1}{p_1-p_2-s} \tag{6.13}$$

Theorem 6.3 *If* $s > c_1 - c_2$*, then* $F(Q_1^*) = \frac{p_1+g-c_1}{p_1+g-s}$ *and* $Q_2^* = 0$.

Proof: We show in this case that it is impossible for both $Q_1^* > 0$ and $Q_2^* > 0$. In particular, if $Q_1^* > 0$ and $Q_2^* > 0$, then KKT conditions (1)

[2]This observation can be seen by rewriting $\frac{p_2+g}{c_2} > \frac{p_1+g-s}{c_1-s}$ as $p_2c_1 - p_1c_2 + g(c_1-c_2) > (p_2+g-c_2)s$.

and (2) imply that $Y_1 = Y_2 = 0$. Hence, equations (6.12) and (6.13) must be satisfied. But if $s > c_1 - c_2$ then from equation (6.13) we would have $F(Q_1^*) > 1$, which would be impossible. Also, we know from Theorem 6.1 that it is impossible for $Q_1^* = 0$ and $Q_2^* > 0$. Hence, $Q_1^* > 0$ and $Q_2^* = 0$. The result follows from KKT condition (3).

Theorem 6.3 states that if the salvage value per unit is greater than $c_1 - c_2$, then the vendor would obtain all items through outright purchase. The purchase amount would be identical to the amount the vendor would purchase if there were no consignment option.

Given this as a background we now turn our attention to how the manufacturer can set the values of p_2 and c_2 to achieve channel coordination. This is examined in the next section.

3. Using Revenue Sharing to Achieve Channel Coordination

The idea of channel coordination is that a manufacturer, through its pricing strategy, will ensure that an independent vendor will order the same amount as if the manufacturer controlled the vendor. In such cases the total expected profit for the channel is maximized.

Channel coordination is desirable to a manufacturer since if total channel profits are maximized while the vendor's expected profit remains constant, the manufacturer's expected profit will be maximized. Of course, one difficulty that can occur with channel coordination is that the vendor's expected profit may increase to such an extent that the manufacturer receives a lower expected profit. In such cases, however, channel coordination can still be desirable to the manufacturer if the manufacturer can obtain a side payment from the vendor equal to a substantial enough portion of the vendor's expected gain.

We will again assume that the vendor has no restrictions on the funding available for obtaining the goods. We make this assumption because a manufacturer would generally not know if a vendor has any limitations on funding and if such limitations existed they would be vendor specific by their nature. Also, if the vendor has limitations on funding, channel coordination may be impossible to achieve. Even if coordination could be achieved, it is doubtful in such cases whether a manufacturer could develop a pricing plan for multiple vendors that would meet the guidelines of the Robinson-Patman act[3].

[3]The Robinson-Patman Act is United States federal legislation that prohibits a manufacturer from engaging in price discrimination among its customers.

We will also assume that the vendor's purchase price and sales price under outright purchase does not change as a result of the manufacturer offering the good through revenue sharing. Instead, channel coordination will be achieved by the manufacturer selecting appropriate values for p_2 and c_2 under a revenue sharing plan. We make this assumption since the manufacturer may have multiple means of distribution, not all of which would be subject to a possible revenue sharing scheme. For example, a video tape manufacturer would sell video tapes not only to video tape rental chains interested in participating in revenue sharing, but also to discount stores, e-commerce retailers, and independent video tape rental stores not wanting to engage in revenue sharing. Our objective is to evaluate the impact that revenue sharing will have on the total channel expected profit as well as the expected profits for both the manufacturer and the vendor.

Let us first focus on the conditions on c_2 and p_2 that will be necessary in order to achieve channel coordination. These are given in the following theorem.

Theorem 6.4 *If the manufacturer wishes to achieve channel coordination then it will be necessary to set* $c_2 = \dfrac{(p_2+g)(m-s)}{p_1+g-s}$.

Proof: Clearly, if the manufacturer wishes to achieve channel coordination, it must set the price for the goods obtained on consignment at an attractive enough level so that $Q_2^* > 0$. We know from Theorem 6.1 that it is impossible for $Q_1^* = 0$ and $Q_2^* > 0$ and from Theorem 6.3 that if $s > c_1 - c_2$ then $Q_2^* = 0$. Hence, c_2 must be set such that $c_1 - c_2 \geq s$ and we will focus on the case where $Q_1^* > 0$ and $Q_2^* > 0$.

If the channel is coordinated, from the results of the standard newsboy model it must be true that Q_1^* and Q_2^* satisfy the following relationship:

$$F(Q_1^* + Q_2^*) \quad = \quad \frac{p_1+g-m}{p_1+g-s} \tag{6.14}$$

Also, we know from KKT condition (4) that:

$$F(Q_1^* + Q_2^*) \quad = \quad \frac{p_2+g-c_2}{p_2+g} \tag{6.15}$$

Equating the two expressions for $F(Q_1^* + Q_2^*)$ results in the relationship for c_2.

Corollary 6.5 *It is necessary to set* $c_2 < (m-s)$ *in order to achieve channel coordination.*

Proof: Equating (6.14) and (6.15) gives:

$$(m-s)\,(p_2+g) = c_2\,(p_1+g-s) \tag{6.16}$$

The result follows since $p_1 - p_2 > c_1 - c_2 \geq s$ and therefore $p_1 - s > p_2$.

The implication of Theorem 6.4 and Corollary 6.5 is that for the manufacturer to achieve channel coordination, it would have to offer consigned goods to the vendor at a cost that is less than the manufacturer's production cost minus the good's salvage value.

Let us assume that the manufacturer selects values of and c_2 and p_2 that satisfy the conditions for achieving channel coordination. In such situations we wish to analyze the effect that achieving channel coordination will have on the vendor's and manufacturer's expected profit.

It should be clear that under channel coordination, the vendor's expected profit will never decline relative to that without channel coordination. This is because the vendor could always adopt a strategy in which $Q_2 = 0$. Hence, while under channel coordination the total expected profit for the manufacturer and vendor increases over the case where the good is not offered on consignment, if the manufacturer does not receive a side payment from the vendor its expected profit might actually decrease. The next theorem gives formulas for the vendor's and manufacturer's expected profit under channel coordination. These formulas can then be used to determine the effect that channel coordination has on the expected profit of the vendor and manufacturer.

Theorem 6.6 *If the manufacturer sets c_2 and p_2 to achieve channel coordination, then:*

$$\begin{aligned} EV(Q_1^*, Q_2^*) &= \int_0^{Q_1^*} (p_1 - s)xf(x)dx \\ &\quad + \int^{Q_1^*+Q_2^*} p_2 xf(x)dx - \int^{\infty} gxf(x)dx \qquad (6.17) \end{aligned}$$

and

$$EM(Q_1^*, Q_2^*) = \int_{Q_1^*}^{Q_1^*+Q_2^*} (p_1 - p_2 - s)xf(x)dx \qquad (6.18)$$

Proof:

$$EV(Q_1^*, Q_2^*) =$$

$$
\begin{aligned}
&= \int_0^{Q_1^*} [p_1 x + s(Q_1^* - x)] f(x)dx + \int_{Q_1^*}^{Q_1^*+Q_2^*} [(p_1 - p_2)Q_1^* + p_2 x] f(x)dx \\
&\quad + \int_{Q_1^*+Q_2^*}^{\infty} [p_1 Q_1^* + p_2 Q_2^* - g(x - Q_1^* - Q_2^*)] f(x)dx - c_1 Q_1^* - c_2 Q_2^* \\
&= Q_1^*(s - p_1 + p_2)F(Q_1^*) - (p_2 + g)(Q_1^* + Q_2^*)F(Q_1^* + Q_2^*) \\
&\quad + \int_0^{Q_1^*} (p_1 - s)x f(x)dx + \int_{Q_1^*}^{Q_1^*+Q_2^*} p_2 x f(x)dx - \int_{Q_1^*+Q_2^*}^{\infty} g x f(x)dx \\
&\quad + p_1 Q_1^* + p_2 Q_2^* + g(Q_1^* + Q_2^*) - c_1 Q_1^* - c_2 Q_2^*. \qquad (6.19)
\end{aligned}
$$

Substituting $F(Q_1^*) = \dfrac{p_1 - p_2 - c_1 + c_2}{p_1 - p_2 - s}$ and $F(Q_1^* + Q_2^*) = \dfrac{p_2 + g - c_2}{p_1 + g}$ into equation (6.19) gives the resulting formula for $EV(Q_1^*, Q_2^*)$.

To show the result for $EM(Q_1^*, Q_2^*)$, we note that:

$$
\begin{aligned}
&EM(Q_1^*, Q_2^*) = \\
&\quad sQ_2^* F(Q_1^*) + \int_{Q_1}^{Q_1^*+Q_2^*} (p_1 - p_2 - s)x f(x)dx \\
&\quad - (p_1 - p_2 - s)Q_1^*[F(Q_1^* + Q_2^*) - F(Q_1^*)] \\
&\quad + sQ_2^*[F(Q_1^* + Q_2^*) - F(Q_1^*)] + (p_1 - p_2)Q_2^*[1 - F(Q_1^* + Q_2^*)] \\
&\quad + (c_1 - m)Q_1^* + (c_2 - m)Q_2^*. \qquad (6.20)
\end{aligned}
$$

Substituting $F(Q_1^*) = \dfrac{p_1 - p_2 - c_1 + c_2}{p_1 - p_2 - s}$ into equation (6.20) gives the following:

$$
\begin{aligned}
EM(Q_1^*, Q_2^*) &= \int_{Q_1}^{Q_1^*+Q_2^*} (p_1 - p_2 - s)x f(x)dx \\
&\quad - (p_1 + g - s)(Q_1^* + Q_2^*)F(Q_1^* + Q_2^*) \\
&\quad + (p_2 + g)(Q_1^* + Q_2^*)F(Q_1^* + Q_2^*) \\
&\quad + (p_1 - p_2 + c_2 - m)(Q_1^* + Q_2^*). \qquad (6.21)
\end{aligned}
$$

Substituting $F(Q_1^* + Q_2^*) = \dfrac{p_1 + g - m}{p_1 + g - s}$ and $F(Q_1^* + Q_2^*) = \dfrac{p_2 + g - c_2}{p_2 + g}$ into equation (6.21) results in the formula for $EM(Q_1^*, Q_2^*)$.

To illustrate the consequences of Theorems 6.4 and 6.6, let us consider the case of a uniform demand distribution, $f(x) = 1/(B-A)$ for $A \leq x \leq B$, 0 elsewhere, and examine the effect that the manufacturer's choice of p_2 has on its expected profit. We select the uniform distribution for demand since it will enable us to generate closed-form expressions for Q_1^*, Q_2^*, and the resulting expected profits for the vendor and the manufacturer.

For a uniform demand distribution the formulas for $EM(Q_1^*, Q_2^*)$ and $EV(Q_1^*, Q_2^*)$ are as follows:

$$\begin{aligned} EV(Q_1^*, Q_2^*) = & \\ & \left[\frac{(p_1 - p_2 - c_1 + c_2)^2}{p_1 - p_2 - s} + \frac{(p_2 + g - c_2)^2}{p_2 + g} - g\right] \frac{(B - A)}{2} \\ & + A\,(p_1 - c) \end{aligned} \tag{6.22}$$

$$\begin{aligned} EM(Q_1^*, Q_2^*) = & \\ & \left[\frac{(p_1 + g - m)(c_1 - m)}{p_1 + g - s} - \frac{(c_1 - m)^2}{2(p_1 - p_2 - s)}\right] (B - A) \\ & + A\,(c_1 - m)\,. \end{aligned} \tag{6.23}$$

(See the appendix for the proof of these two expressions.)

We therefore see from equation (6.23) that the manufacturer's expected profit is monotonically decreasing in p_2 and the manufacturer's optimal strategy would be to set p_2 as low as possible. Unfortunately, if p_2 is restricted to being nonnegative, it may not be possible for the manufacturer to set p_2 low enough to ensure that its expected profit will increase over the case in which revenue sharing is not offered. The conditions required for the manufacturer's profit to not decline under channel coordination are given in the next theorem.

Theorem 6.7 *If demand follows a uniform distribution and the manufacturer wishes to set consignment pricing to achieve channel coordination, the manufacturer will need to set c_2 to be less than or equal to $(m - s)/2$ in order for its expected profits to not decline.*

Proof: Without channel coordination, the manufacturer's expected profit will be:

$$EM(Q_1^*) = \left(\frac{p_1 + g - c_1}{p_1 + g - s}\right)(c_1 - m)(B - A) + A\,(c_1 - m) \tag{6.24}$$

Comparing equation (6.24) with equation (6.23) gives the desired result.

Corollary 6.8 *If demand follows a uniform distribution and the manufacturer sets consignment pricing to achieve channel coordination so that its expected profits do not decline, a necessary condition for* $p_2 \geq 0$ *is that* $p_1 \geq g + s$.

Proof: The result follows from Theorems 6.4 and 6.7.

Examples

To illustrate these concepts, let us consider the following two examples. In the first example we deal with a case in which $p_1 \leq g + s$. In particular, suppose:

$$p_1 = \$100, c_1 = \$40, s = \$10, g = \$100, m = \$20, \text{ and } f(x) = U(0, 95).$$

If the vendor does not have the option of obtaining items through consignment, we know from the newsboy problem solution that the vendor's optimal purchase quantity would be $Q_1^* = 80$. In this case the vendor would earn an expected profit of:

$$\begin{aligned} EV(80) &= \int_0^{80} [100x + 10(80 - x)] \frac{1}{95} dx \\ &\quad + \int_{80}^{95} [100 * 80 - 100(x - 80)] \frac{1}{95} dx - 40 \cdot 80 \\ &= \$1,650 \end{aligned}$$

and from equation (6.24) we have that the manufacturer would earn a profit of 80 • ($40 – $20) = $1,600. Total channel expected profit in this case would therefore be $3,250.

Under channel coordination, the optimal order quantity would satisfy the relationship: $F(Q^*) = \dfrac{p_1 + g - m}{p_1 + g - s}$, which for this example, results in a value of $Q^* = 90$. The total channel expected profit in this case would be:

$$\begin{aligned} EP(90) &= \int_0^{90} [100x + 10(90 - x)] \frac{1}{95} dx \\ &\quad + \int_{90}^{95} [100 \cdot 90 - 100(x - 90)] \frac{1}{95} dx - 20 \cdot 90 \\ &= \$3,350. \end{aligned}$$

Thus, by coordinating the channel, total channel expected profits would increase by \$100 or 3.1%.

Now, let us assume that the manufacturer offers the good on consignment to the vendor at a cost of \$8, e.g. $c_2 = 8$. We know from Theorem 6.4 that p_2 should be set to satisfy the relationship:

$$p_2 = \frac{c_2(p_1 + g - s)}{m - s} - g.$$ Hence, p_2 should be set equal to 52.

In this case, we can use equations (6.12) and (6.13) to determine the values of Q_1^* and Q_2^*. Solving the relationships:

$$F(Q_1^* + Q_2^*) = 0.94737 \text{ and } F(Q_1^*) = 0.42105$$

gives $Q_1^* = 40$ and $Q_2^* = 50$.

For these values the vendor's expected profit would be:

$$\begin{aligned} EV(40, 50) &= \\ &= \left[\frac{(100 - 52 - 40 + 8)^2}{100 - 52 - 10} + \frac{(52 + 100 - 8)^2}{52 + 100} - 100\right]\frac{(95 - 0)}{2} \\ &= \$2,050 \end{aligned}$$

and the manufacturer's expected profit would be:

$$\begin{aligned} EM(40, 50) &= \\ &= \left[\frac{(100 + 100 - 20)(40 - 20)}{(100 + 100 - 10)} - \frac{(40 - 20)^2}{2(100 - 52 - 10)}\right](95 - 0) \\ &= \$1,300. \end{aligned}$$

Thus, we see that even though channel coordination is achieved, this is at the expense of the manufacturer seeing a \$300 decline in expected profit.

In fact, for this example, unless the manufacturer selects a value of c_2 less than or equal to \$5, the manufacturer's expected profit would be less than if it did not offer the goods on consignment. If, however, $c_2 \leq \$5$, then $p_2 \leq -\$5$. Since a manufacturer would have great difficulty convincing a vendor to take goods on consignment if $p_2 < 0$, revenue sharing in this case would result in a lower expected profit to the manufacturer. As mentioned earlier, one possible way around this dilemma would be for the manufacturer to require the vendor to make a side payment to the manufacturer. This side payment would have to be equal to a large enough portion of the vendor's expected increase in profit so that both the vendor and manufacturer show some gain from revenue sharing. While such an approach has appeal, the manufacturer may also face difficulty in convincing the vendor to make such a side payment.

Now let us consider a case in which $p_1 \geq g+s$. In particular, suppose:

$$p_1 = \$100, c_1 = \$40, s = \$10, g = \$50, m = \$20, \text{ and } f(x) = U(0, 98).$$

In this case, without revenue sharing the vendor's optimal order quantity would be $Q_1^* = 77$, the vendor's expected profit would be \$1,785, and the resulting profit to the manufacturer would be \$1,540. If the channel were coordinated, the total channel expected profit would be \$3,465, representing a potential increase of \$140 or 4.2%. If the manufacturer sets $c_2 = \$4$, for example, then p_2 would need to be set at \$6 in order to achieve channel coordination. Using these values, the manufacturer's expected profit would increase to \$1586.67 and the vendor's expected profit would increase to \$1,878.33. Here we see that since $p_1 \geq g+s$, channel coordination can be achieved in a manner that increases both the vendor's and the manufacturer's expected profits.

4. Conclusion

As we have shown, revenue sharing is an intriguing method for a manufacturer to achieve channel coordination. Unfortunately, if the manufacturer wishes to maintain the current pricing structure for the good while allowing for revenue sharing, it is possible that without a side payment from the vendor the manufacturer's expected profit would actually decrease.

There are several avenues for further research in this area. In particular, one could investigate the situation in which the manufacturer is free to change both c_1 and c_2 and the vendor is free to set p_1. Such an analysis would require an estimation of the demand curve. Another possible avenue for further research is to extend the results of Section 3 to distributions other than the uniform. A third possible extension of the model is to a multi-vendor environment and to determine the pricing policy that maximizes the total channel profit in such situations.

Appendix

Proof of Equation (6.22): From Equation (6.17) we have for a uniform distribution that:

$$EV(Q_1^*, Q_2^*) = \left.\frac{(p_1 - s)x^2}{2(B-A)}\right|_A^{Q_1^*} + \left.\frac{(p_2 x^2}{2(B-A)}\right|_{Q_1^*}^{Q_1^*+Q_2^*} - \left.\frac{gx^2}{(B-A)}\right|_{Q_1^*+Q_2^*}^{B} \qquad (6.A.1)$$

From Equation (6.13) we have that

$$Q_1^* = (B - A) \quad \frac{p_1 - p_2 + c_2 - c_1}{p_1 - p_2 - s} \quad + A \tag{6.A.2}$$

while from Equation (6.15) we have that

$$Q_1^* + Q_2^* = (B - A) \quad \frac{p_2 + g - c_2}{p_2 + g} \quad + A \tag{6.A.3}$$

Substituting (6.A.2) and (6.A.3) into (6.A.1) and simplifying gives

$$EV(Q_1^*, Q_2^*) = \frac{(p_1 - p_2 - c_1 + c_2)^2}{p_1 - p_2 - s} + \frac{(p_2 + g - c_2)^2}{p_2 + g} - g \quad \frac{(B - A)}{2} + A\,(p_1 - c)\,. \tag{6.A.4}$$

Proof of Equation (6.23): From Equation (6.18) we have for a uniform distribution that

$$EM(Q_1^*, Q_2^*) = \frac{(p_1 - p_2 - s)x^2}{2(B - A)} \Big|_{Q_1^*}^{Q_1^* + Q_2^*} \tag{6.A.5}$$

or

$$EM(Q_1^*, Q_2^*) = \frac{(p_1 - p_2 - s)(2Q_1^* + Q_2^*)Q_2^*}{2(B - A)}. \tag{6.A.6}$$

From Equation (6.14) we have

$$Q_1^* + Q_2^* = (B - A) \quad \frac{p_1 + g - m}{p_1 + g - s} \quad + A. \tag{6.A.7}$$

Define $Q_1^* = \alpha(B - A) + A$, where $\alpha = \frac{p_1 - p_2 + c_2 - c_1}{p_1 - p_2 - s}$ and $Q_1^* + Q_2^* = \beta(B - A) + A$, where $\beta = \frac{p_2 + g - c_2}{p_2 + g}$. Then from (6.A.6) we have:

$$EM(Q_1^*, Q_2^*) = (p_1 - p_2 - s)(\beta - \alpha) \quad A + \frac{(\alpha + \beta)(B - A)}{2} \quad . \tag{6.A.8}$$

But, from the definition of α and β we have:

$$\begin{aligned}
(p_1 - p_2 - s)(\beta - \alpha) &= (p_1 - p_2 - s) \quad \frac{(p_1 + g - m)}{(p_1 + g - s)} - \frac{(p_1 - p_2 + c_2 - c_1)}{(p_1 - p_2 - s)} \\
&= \frac{(p_1 + g - s + s - m)(p_1 - p_2 - s)}{(p_1 + g - s)} - p_1 + p_2 - c_2 + c_1 \\
&= \frac{(s - m)(p_1 + g - s - p_2 - g)}{(p_1 + g - s)} + c_1 - c_2 - s \\
&= \frac{(m - s)(p_2 + g)}{p_1 + g - s} + c_1 - c_2 - m \\
&= c_1 - m.
\end{aligned} \tag{6.A.9}$$

The last substitution in (6.A.9) follows from Equation (6.16). From (6.A.9) we have that

$$\begin{aligned}\frac{p_1 - p_2 - s}{2}\left[(\beta - \alpha)(\beta + \alpha)\right] &= \frac{(c_1 - m)(\beta + \alpha)}{2} \\ &= \frac{(c_1 - m)(\alpha - \beta + 2\beta)}{2}. \qquad (6.A.10)\end{aligned}$$

But from (6.A.9) we have that $(\alpha - \beta) = -\frac{(c_1 - m)}{p_1 - p_2 - s}$. Hence, from (6.A.9) and (6.A.10) we have

$$EM(Q_1^*, Q_2^*) = -\frac{(c_1 - m)^2}{p_1 - p_2 - s} + \frac{2(c_1 - m)(p_1 + g - m)}{p_1 + g - s} \quad \frac{B - A}{2} \quad + A(c_1 - m). \qquad (6.A.11)$$

Rearranging the terms in (6.A.11) gives Equation (6.23)

References

Brown, A. O. and Lee, H.L. 1998, "The Win-Win Nature of Options-Based Capacity Reservations Arrangements," Owen School Working Paper.

Drezner, Z and Pasternack, B.A. 1999, "The Videotape Rental Model," *Journal of Applied Mathematics & Decision Sciences,* 3, 167-174.

Hillier, F. and Lieberman, G. 2001, *Introduction to Operations Research,* New York, McGraw-Hill Higher Education.

Jeuland, A. P. and Shugan, S. M. 1983, "Managing Channel Profits," *Marketing Science,* 2, 239-272.

Jucker, J. V. and Rosenblatt, M. J. 1985, "Single-Period Inventory Models with Demand Uncertainty and Quantity Discounts: Behavioral Implications and a New Solution Procedure," *Naval Research Logistics Quarterly,* 32, 537-550.

Kaback, I. and Weinberg, C. 1972, "The Generalized Newsboy Problem, Contract Negotiations, and Secondary Vendors," *IIE Transactions,* 4, 154-157.

Khouja, M. 1996, "The Newsboy Problem with Progressive Retailer Discounts and Supplier Quantity Discounts," *Decision Science,* 27, 589-599.

Khouja, M. 1999, "The Single-Period (News-Vendor) Problem: Literature Review and Suggestions for Future Research," unpublished research report.

Lal, R. 1990, "Improving Channel Coordination Through Franchising," *Marketing Science,* 9, 299-318.

Lau, H. and Lau A. H. 1999, "Manufacturer's Pricing Strategy and Return Policy for a Single-Period Commodity," *European Journal of Operational Research,* 116, 291-304.

Lawrence, J. A. and Pasternack, B. A. 1998, *Applied Management Science, A Computer Integrated Approach for Decision Making,* New York, John Wiley & Sons.

Lariviere, M. A. and Porteus, E. L. 1999, "Selling to the Newsvendor," Duke University Working Paper.

Lin, C. and Kroll, D. E. 1997, "The Single-Item Newsboy Problem with Dual Performance Measures and Quantity Discounts," *European Journal of Operational Research,* 100, 562-565.

Monahan, J. P. 1984, "A Quantity Discount Pricing Model to Increase Vendor Profits," *Management Science,* 30, 720-726.

Padmanabhan, V. and Png, I. P. L. 1997, "Manufacturer's Returns Policies and Retail Competition," *Marketing Science,* 16, 81-94.

Pasternack, B. A. 1985, "Optimal Pricing and Return Policies for Perishable Commodities," *Marketing Science,* 4, 166-176.

Pasternack, B. A. 2000 "The Capacitated Newsboy Problem with Revenue Sharing," *Journal of Applied Mathematics & Decision Sciences,* 5, 7-29.

Silver, E.A., Pyke, D.F., and Peterson, R. P. 1998, *Inventory Management and Production Planning and Scheduling,* Third Edition, New York, John Wiley & Sons.

Tsay, A.A., Nahmias, S., and Agrawal, N. 1999, "Modeling Supply Chain Contracts: A Review," *Quantitative Models for Supply Chain Management,* (Tayur, S., R. Ganeshan, and M. Magazine, eds.), Kluwer Academic Publishers.

Chapter 7

SUPPLY CHAIN CONTRACTING AND COORDINATION WITH SHELF-SPACE-DEPENDENT DEMAND

Yunzeng Wang

Department of Operations, Weatherhead School of Management
Case Western Reserve University
Cleveland, OH 44106
yxw36@po.cwru.edu

Yigal Gerchak

Department of Industrial Engineering,
Tel-Aviv University
Tel-Aviv 69978
Israel
ygerchak@eng.tau.ac.il

Abstract Consider a manufacturer or wholesaler who supplies some item to retailers facing demand rates which depend on the shelf or display space devoted to that product by themselves and their competitors. The manufacturer, via the use of financial levers at her disposal, wishes to coordinate this decentralized chain while making a profit. With a single retailer, we show that the manufacturer can achieve this goal by using a two-parameter contract: a wholesale price and an inventory holding costs subsidy offered to the retailer. When multiple retailers compete in that product's market, there are two ways to envision and model the demand and market split. One assumes that market demand depends on aggregate inventory displayed, and then splits according to individual display levels. The other "assigns" customers to retailers according to their display levels, and then assumes that purchases are a function of the display level at the retailer selected. We characterize retailers' Nash equilibria in these models, and explore whether the manufacturer can coordinate such channels. Information requirements for channel co-

J. Geunes et al. (eds.),
Supply Chain Management: Models, Applications, and Research Directions, 137—162.

ordination and profit allocation are discussed throughout the analysis.

1. Introduction

Advertising, promotions, extended business hours, improved service and product variety are well known means for attracting more demand (e.g., Eliashberg and Lilien 1993 and references therein). Retailers can often affect sales volume of a product or a product family by increasing the shelf space allocated to it. It is recognized by marketing/consumer behavior researchers and practitioners that, with certain types of items, the quantity displayed can have a motivational effect on demand and sales (Schary and Becker 1972, Kollat and Willet 1969, Krugman 1965 and Corstjens and Doyle 1981). At least two types of stimuli effects of inventory on demand have been identified. The first and more obvious one is referred to as the "selective effect", where more items in inventory provide customers with more to choose from and thus induce them to purchase or purchase more. This will happen where the units of an item (e.g., fresh fruits, vegetables, etc.) are not identical, and a customer may like the feeling of a wide "selection". For certain baked goods (e.g., doughnuts), low stocks may raise customers' perception that the units are "left-overs" and not fresh (Pasternack 1990 refers to this phenomenon as "balking").

The second type of stimulus of inventory is the "advertising effect". A large displayed quantity often gives consumers the impression that the item is *popular* in the market, which may signal good "value" (low price, high quality, etc.) and hence induce more consumers to buy it and each to buy more. Thus, displayed inventory acts as a way of advertisement.

For these reasons, manufacturers often like their products to be widely displayed, and retailers want their shelf space fully utilized/filled. Since increased shelf space or displays often require the retailer to keep higher inventories, choosing an item's shelf space is part of choosing its inventory level. Thus, some operations management researchers have incorporated inventory-level-dependence of demand into various inventory control models (Johnson 1968, Iglehart and Jaquette 1969, Baker and Urban 1988). Wang (1992, Chapter 5) and Parlar et al. (1994) considered general EOQ models (possibly with random yield) with inventory-level-dependent demand rates. Gerchak and Wang (1994) analyzed a periodic-review model where the demand per period is the product of a random variable and a function that increases in the period's inventory level. This literature dealt exclusively with a single decision-maker (the retailer).

Since a manufacturer's or wholesaler's decisions (e.g., on wholesale prices) affect their retailers' inventory policies, it seems important to take into account the dependence of demand on the displayed inventory level, and to address issues related to supply chain coordination in such environments. So far, however, most of the coordination literature (e.g., Jeuland and Shugan 1983, Eliashberg and Steinberg 1993, Tayur et al. 1999) has viewed demand, whether deterministic or stochastic, as exogenous and unaffected by the chain's activities (other than the retail price). The model by Tsay and Agrawal (1998; see also their references) does include demand dependence on service as well as price (in a competitive setting), which is related to our "exposure" concept, but their linear demand function is quite different from our market split models. It is our goal to analyze a decentralized manufacturer-retailer(s) supply chain, which, though physically greatly simplified, recognizes the positive dependence of demand on the quantity displayed.

The specific supply chain we try to model can be envisioned as a marketplace of multiple vendors/retailers who are located in close proximity, perhaps within the same open area or the same building. These types of marketplaces are widespread and best exemplified by fresh food (fruits, vegetables, fish, meats, etc.) vendors who can easily be found in any Chinatown district, Italian market or farmers' market. These vendors' operations can be characterized by the following common features. Vendors in the market often stock/sell the same product(s) (e.g., apples, lettuce, etc.), which are often purchased from the same grower/distributor. Second, since shelf space is very costly/limited, a vendor replenishes her shelf essentially as soon as a customer withdraws inventory from the shelf. (Since the shelf space is often very limited, say, a few dozen square feet, "reviewing" the stock level continuously and reacting to changes immediately is very easy.) Third, customers can readily observe the displayed inventory of all the vendors in the marketplace, since stalls are very close to, or even attached to, each other. Also, customers often like to walk through the market before selecting which vendor to buy from.

We first analyze a stylized model by assuming that a supplier ("manufacturer") supplies some product to a single retailer. The retailer decides to allocate some shelf space to display S units of a product. Once allocated, the shelf is replenished via the $(S-1,S)$ policy with zero lead time, so the displayed inventory will be kept at level S all the time. That mimics store replenishment situations when replenishments are very frequent. The demand rate depends on S via a general increasing concave function. The item's retail price is assumed to be fixed due to competition, and the inventory holding (shelf space) cost rate is constant (linear). With a fixed wholesale price charged by the manufacturer, the displayed amount that maximizes the retailer's long-term average profit can be uniquely determined from her demand function and cost parameters.

The manufacturer's decisions in a decentralized setting pertain to the financial arrangement with the retailer. The key variable is, of course, the wholesale price charged. If the manufacturer chooses to coordinate the chain using only the wholesale price – a price-only contract – its profit will be zero, since the only such contract which coordinates the channel is for the wholesale price to be equal to the manufacturer's marginal production cost. (We follow the literature in assuming that the manufacturer cannot appropriate all the rents by charging the retailer a fixed fee.) If the manufacturer maximizes its profit, the displayed level (and sales) by the retailer will be lower than the system-optimal level, and thus the total-profit pie smaller, due to the well-known "double marginalization" phenomenon (Spengler 1950). We provide an example of a situation where the manufacturer needs no information about the retailer's holding cost to set up the optimal wholesale-price, and another example where he does need this information.

To coordinate the channel as well as to make a profit, the manufacturer needs an additional financial lever. We suggest a holding cost subsidy offered to the retailer, which will cause the retailer to stock more. With the two levers – the price-plus-inventory-subsidy contract, the manufacturer can achieve not only channel coordination, but also any desired allocation of the channel profit between himself and the retailer. We show that as far as channel coordination is concerned, the manufacturer does *not* need to know the inventory-level-dependent demand function to offer such a contract. But, for the purpose of channel profit allocation, the manufacturer may need the demand information, depending on whether a percentage or a dollar-amount type of allocation is to be achieved. We also generalize the scenario to non-linear (convex) inventory holding costs and show that a price-plus-inventory-subsidy contract can still coordinate the supply chain. On the other hand, compared with the linear cost case, the design of a coordinating contract will require the manufacturer to possess information about the retailer's demand function.

We then extend our analysis to two retailers who share, and compete in, the same market of the supply chain. Displayed inventory by a retailer now has two functions: to motivate customer demand, and to gain more market share. We propose two ways to model the customer demand and retailers' competition processes. The first model envisions a situation where a customer selects her demand quantity based on the *aggregate* inventory of both retailers and then chooses from which retailer to buy based on their relative inventory levels. This aggregate-inventory dependent demand function is motivated by the "advertising effect" of inventory on demand as discussed earlier. The basic assumption is that a customer selects her consumption based mainly on her perception of the item's "popularity", and the total displayed inventory she sees in the market is the major determinant of her perception. As described earlier, a customer can indeed observe the total inventory of the "whole market" (i.e., of all retailers) in the type of marketplace we intend to model.

In the second model, a customer will first choose the retailer based on the relative amounts of inventories displayed by both retailers and then decide on the purchase quantity based on *individual* inventory level at the chosen retailer. This demand function form applies to situations where the "selective effect" of inventory is the predominant factor determining a customer's consumption quantity. The individual inventory of the retailer chosen by the customer will *limit/provide* the "pool" of units from which she can select the ones she likes, assuming that the transaction cost for her to buy the same item from both retailers will be too high. Of course, any real-world situation or a customer's decision process will be much more complex than any of these two *stylised* models can capture.

How does a customer choose from which retailer to buy? We assume for both models that the split of customers between the two retailers is proportional to their relative inventory. The rationale here is that a customer would prefer to buy from a bigger pool of inventory, which is supported by the "selective effect" of inventory. Since displayed inventory acts as means of marketing/advertising efforts, this proportional market split form is also consistent with the general market share models where the determinants correspond to advertising (Moorthy 1993, Section 5.1) and to any general marketing efforts (Cooper 1993, p. 262; Monahan 1987). In fact, Kotler (1984, p. 231) refers to such market split as the "Fundamental Theorem of Market Share". In contrast, Deneckere and Peck (1995), Dana and Spier (2000) and Dana (2000) envision a situation where the total (random) demand is not influenced by a firm's inventories (capacities), but where customers select firms so as to maximize their chances of obtaining the product.

Concrete insights into duopoly equilibrium and supply chain coordination are then obtained by assuming that each customer's demand is a concave power function of the displayed inventory. Specifically, for the first model of competition (where demand is a function of aggregate inventory), we show that there exists a unique Nash equilibrium of displayed inventories for any given contract offered by the upstream manufacturer. As expected, two competing retailers always display more inventory in total than a monopolist retailer for a given manufacturer's contract, and thus competition would generate inefficiency at the downstream of the supply chain. Surprisingly, however, the manufacturer, by using the price-plus-inventory-subsidy contract, can still achieve channel coordination for this competitive supply chain! On the other hand, compared with the single retailer chain, coordination here will require the manufacturer to know and use the demand-function information.

It turns out that the second model of competitive retailers (where demand is a function of individual inventory) is much more complex to analyze. We characterize one symmetric equilibrium solution of the problem. One key insight we obtain here is that, with demand function of individual inventory, the manufacturer cannot coordinate a competitive supply chain. As we show, for the first model, the "efficiency" of the supply chain depends only on the *aggregate* (total) displayed inventory (but not on the allocation of inventory between the two retailers), which the manufacturer can control through contract parameters. For the second model, however, supply chain efficiency depends not only on the aggregate inventory but also on the inventory allocation between the two retailers. Inventory allocation is determined through a competition mechanism, over which the manufacturer has no control. Specifically, for the second model we show that splitting inventory among two retailers would cause waste/inefficiency. Thus, any attempt (or contract type) to coordinate the channel will require the contingency of closing down one of the retailers.

Although our models have been motivated by the concrete settings of fresh-food/farmers-markets, the model assumptions can be relevant to many other, if not all, retailing industries, like automobile dealerships and supermarkets. As mentioned earlier, the motivational effects of displayed inventory on demand/sales are widely recognized by practitioners and discussed by marketing and consumer-behavior researchers. Balakrishnan et al. (2000) report on stores which display huge inventories (sufficient to meet many years of demand) of items like blank videotapes, clearly aiming at stimulating demand. Thus, our model analyses and managerial insights generated can well be relevant to the retailing industry in general. It is also interesting to observe that the booming Internet is a marketplace where all retailers are located virtually in the same "place", and many of the Internet retailers also provide information on stock levels. These features actually fit well into out model settings.

Finally, we note that our work is not the first to suggest inventory-holding subsidies to downstream retailers in decentralized supply chains.

Anupindi and Bassok (1999) use such an incentive mechanism to design decentralized distribution systems and to identify conditions that will benefit the different parties involved. Moses and Seshadri (1996) propose extended credit-terms, which act like holding cost subsidies. However, none of these studies addresses issues related to channel coordination.

The rest of this paper is organized as follows. In Section 2, we present our model and analysis for a manufacturer-single-retailer supply chain, centralized and decentralized. Sections 3 and 4 study the manufacturer-competitive-retailers models with demand being a function of aggregate and individual inventory, respectively. We then make some concluding remarks in Section 5. For the convenience of readers, all mathematical proofs are relegated to an Appendix.

2. A Manufacturer and Single Retailer Supply Chain

2.1 The Model and Centralized Control

A single product is produced by a manufacturer and then sold to consumers through (for now) a single retailer. The marginal production cost and retail price are constant at c \$/unit and p \$/unit respectively, where $p > c$. The demand rate for the item will depend on the amount of inventory displayed at the retailer's shelf. Specifically, a constant inventory level of I units generates a demand of $D(I)$ units/year. In general, $D(I)$ can be assumed to be an increasing and concave function (i.e., $D'(I) > 0$ and $D(I) < 0$) to reflect the motivational effect of inventory on demand and the "diminishing returns". For technical purposes, we also assume that $D(I)$ is continuous and twice differentiable with $D'(0) \to \infty$, $D'(\infty) \to 0$ and $D(I) > 0$ for any $I > 0$. Displaying inventory at the retailer is costly. Assume that a constant inventory cost of h \$/unit/year is charged. So the key decision here is to choose the displayed inventory level I to trade-off increased sales against inventory costs. Note that, once I is chosen, the system is assumed to keep the inventory at level I all the time by continuously replenishing it.

Now, if this supply chain is centrally owned and controlled, the objective is to maximize the long-run average channel profit (i.e., the profit rate):

$$\underset{I \geq 0}{Max}\, \Pi^c(I) = (p-c)D(I) - hI, \tag{7.1}$$

where the first term is the sales revenue net of production cost and the second term is the inventory holding cost. One can easily verify that $\Pi^c(I)$ is concave and thus the unique solution I^c is given by the first order condition:

$$D'(I^c) = h/(p-c) \quad \text{i.e.,} \quad I^c = D'^{-1}(h/(p-c)). \qquad (7.2)$$

We next consider a decentralized system where the manufacturer, through contractual arrangements, wholesales the product to the retailer who then chooses its displayed inventory level and, hence, the demand rate. We consider two specific contractual arrangements: a price-only contract, in which the manufacturer selects only the wholesale price, and a price-plus-inventory-subsidy contract where a wholesale price plus an inventory cost subsidy are offered by the manufacturer.

2.2 Price-Only Contract

Here, the manufacturer offers the retailer a take-it-or-leave-it contract which specifies only a wholesale price, say, w \$/unit. If the retailer takes the contract, she then selects a (permanent) displayed inventory level I which determines the sales rate $D(I)$. For simplicity, assume that both the manufacturer and the retailer have an opportunity cost of zero. So, as long as w is chosen such that $c < w < p$, it will be a viable contract for both parties. In determining the inventory level, the retailer wishes to maximize her own profit (rate):

$$\underset{I \geq 0}{Max}\, \Pi^r(I) = (p-w)D(I) - hI. \qquad (7.3)$$

The unique optimal inventory level I^r for the retailer is thus given by

$$D'(I^r) = h/(p-w) \quad \text{i.e.,} \quad I^r = D'^{-1}(h/(p-w)). \qquad (7.4)$$

Since $D'(I)$ is a decreasing function, by simply comparing (7.4) with (7.2), we see that, as long as the manufacturer charges a wholesale price w above his marginal cost c, the inventory level I^r in a price-only contractual arrangement will always be lower than the inventory level I^c in a centralized system. This phenomenon is essentially the "double marginalization" problem studied in the economics and industrial organization literature (Spengler 1950, Cachon 1999 and Lariviere 1999). We assume that the manufacturer is not able to charge the retailer a fixed fee. If that was feasible, then by setting $w = c$ the manufacturer could coordinate the system and extract all the channel profit.

Knowing that the retailer will choose the inventory level according to (7.4), the manufacturer chooses the wholesale price w so as to maximize his own profit:

$$\underset{c<w<p}{Max}\, \Pi^m(w) = (w-c)D(I^r). \qquad (7.5)$$

Example 7.1 *Let* $D(I) = aI^b$, $0 < b < 1$. *Then,* $D'(I) = abI^{b-1}$, *so* $D'^{-1}(y) = ab/y^{1/(1-b)}$. *Thus,*

$$\begin{aligned}\Pi^m(w) &= (w-c)a\left[\left(\frac{ab}{h/h(p-w)}\right)^{1/(1-b)}\right]^b, \\ &= h^{(1-b)/b}a^{1/(1-b)}b^{b/(1-b)}(w-c)(p-w)^{b/(1-b)}.\end{aligned}$$

We observe that the optimal value of w *does* not *depend on* h. *Thus, whether or not the manufacturer has information about retailer's holding cost, the optimal wholesale price is not affected. Maximizing* $\Pi^m(w)$ *in this example amounts to maximizing*

$$(w-c)(p-w)^{b/(1-b)} \equiv A(w).$$

Now, $A'(w) = (p-w)^{b/(1-b)}\left(1 - \frac{w-c}{p-w}\right)$.
For $c < w < p$, *solving* $A'(w) = 0$ *for* w, *we get the* unique *solution*

$$w* = \frac{p+c}{2}.$$

Furthermore, $A'(w) > 0$ *as* $w \to c$ *and* $A'(w) < 0$ *as* $w \to p$. *Thus,* $w*$ *is the unique maximizer of* $A(w)$ *and, hence, of* $\Pi^m(w)$.

Example 7.2 *Let* $D(I) = a\ln(b+I)$, $a, b > 0$. *Here,* $D'(I) = a/(b+I)$, *so* $D'^{-1}(y) = a/y - b$. *Thus,*

$$\begin{aligned}\Pi^m(w) &= (w-c)a\ln\left\{b + \left[\frac{a}{h/(p-w)} - b\right]\right\} \\ &= a(w-c)\left[\ln a + \ln(p-w) - \ln h\right] \equiv B(w).\end{aligned}$$

We then have $B'(w) = a\left[\ln a + \ln(p-w) - \ln h - \frac{w-c}{p-w}\right]$, *and*

$$B''(w) = a\left[-\frac{1}{p-w} - \frac{p-c}{(p-w)^2}\right] < 0.$$

The optimality condition is then

$$\ln(p-w) - \frac{w-c}{p-w} = \ln(a/h).$$

So, here w *does depend on* h.

For the general optimization problem of (7.5), the concavity of $\Pi^m(w)$ is not guaranteed for all demand function forms $D(\cdot)$. To generate some further insights into the uniqueness of the solution to (7.5), we note that there is a one-to-one correspondence between w and I^r – they are related to each other through (7.4). So, when the manufacturer chooses a value for w $(c < w < p)$, he is equivalently choosing a value for I^r $(I^c > I^r > 0)$. Thus, substituting $w = p - h/D'(I^r)$, the optimization problem over w in (7.5) can be written as the following optimization over I^r:

$$\underset{I^c > I^r > 0}{Max}\, \Pi^m(I^r) = [p - h/D'(I^r) - c]D(I^r). \tag{7.6}$$

(A similar approach was employed by Lariviere and Porteus 1999).

Taking the derivative of $\Pi^m(I^r)$, we have,

$$\frac{d\Pi^m(I^r)}{dI^r} = hD''(I^r)D(I^r)/[D'(I^r)]^2 + [p - h/D'(I^r) - c]D'(I^r). \tag{7.7}$$

Now, the second term in (7.7) is always positive and decreasing and the first always negative. Thus, for a given demand function $D(\cdot)$, if the first term in (7.7), i.e., $D''(I^r)D(I^r)/[D'(I^r)]^2$ is non-increasing, then the manufacturer's profit function is unimodal and, hence, the solution to (7.6) is unique. For the demand functions of Examples 7.1 and 7.2, one can check that this condition is always satisfied.

2.3 Price-Plus-Inventory-Subsidy Contract

Suppose now that the manufacturer offers the retailer a wholesale price of w \$/unit and an inventory holding subsidy of s \$/unit/year towards any inventory the retailer chooses to hold on shelf. The retailer's problem then becomes

$$\underset{I \geq 0}{Max}\, \Pi^r(I) = (p - w)D(I) - (h - s)I, \tag{7.8}$$

and her optimal inventory level I^r is given by

$$D'(I^r) = (h - s)/(p - w) \text{ i.e., } I^r = D'^{-1}\left((h - s)/(p - w)\right). \tag{7.9}$$

Comparing (7.9) with (7.2), we have the following important observation:

Proposition 7.3 *For any contract* (w, s) *such that*

$$(h - s)/(p - w) = h/(p - c), \text{ i.e., } s = h - h(p - w)/(p - c), \tag{7.10}$$

we have $I^r = I^c$.

In other words, for *any* wholesale price w, $c < w < p$, offered by the manufacturer, if he accordingly chooses an inventory subsidy $s =$

$h - h(p-w)/(p-c)$, then the retailer will always be induced to choose the centralized-system-optimal inventory level I^c, and, hence, such a contract *coordinates* the decentralized supply chain. Thus, there exists a continuum of (w, s) contracts that coordinate the supply chain.

In choosing (w, s), the manufacturer will want to maximize his own profit. So, presumably, he solves the following problem:

$$\underset{c \leq w \leq p, s \geq 0}{Max} \Pi^m(w, s) = (w - c)D(I^r) - sI^r, \tag{7.11}$$

where I^r is determined through (7.9). The first term in (7.11) is his sales revenue net of production cost, and the second term is her inventory subsidy to the retailer.

Instead of solving (7.11) directly, the following argument (Pasternack 1985) illustrates how the manufacturer can find his best strategy: Focus on the set of contracts (w, s) which satisfy (7.10) with $c \leq w \leq p$. We know from Proposition 7.3 that any contract within this set will coordinate the channel and, hence, achieve the maximum channel profit. But, as we will show next, different contracts within this set, represented by different values of w, provide the retailer with a different amount of profit - the rest of the maximal channel profit will go to the manufacturer. As a consequence, the manufacturer needs simply to choose a value of w so as to allocate any amount of profit required by the retailer (so that she will accept the contract) and thus to extract as much profit out of the supply chain as he can! Now, with $s = h - h(p-w)/(p-c)$ and $I = I^c$, after some algebra, the retailer's profit in (7.8) can be written as

$$\Pi^r(I^c) = [D(I^c) - hI^c/(p-c)](p-w), \tag{7.12}$$

which is linearly decreasing in w.

The total channel profit $\Pi^c(I^c)$ can be obtained simply by substituting $I = I^c$ into (7.1). We can show that

$$\Pi^r(I^c)/\Pi^c(I^c) = (p-w)/(p-c). \tag{7.13}$$

So, the proportion of the channel profit allocated to the retailer is also linearly decreasing in w.

To summarize, a properly designed price-plus-inventory-subsidy contractual arrangement (w, s) can achieve: 1. coordination of the supply chain channel; 2. any desired allocation of channel profit between the manufacturer and the retailer.

Finally, we discuss the information the manufacturer will need in order to coordinate this supply chain. First, we see from (7.10) that for any given wholesale price, w, the manufacturer only needs the cost parameters (i.e., p, c and h) to determine a corresponding inventory subsidy

s. Here he does not need information about the demand function $D(\cdot)$. The value of w will determine the portion/amount of channel profit allocated to the retailer. If he is only interested in the proportion allocated, then equation (7.13) shows that the manufacturer still does *not* need the demand information. Only if his aim is to achieve a specific allocation of absolute profits does he need this information - equation (7.12).

2.4 Non-Linear Holding Costs

We now relax the linear holding/shelf space cost assumption by considering a general convex cost function, which is arguably more realistic for most retailing situations where shelf space is a limited resource. We derive the retailer's optimal inventory decision and discuss if and how the manufacturer can coordinate the supply chain.

Assume that, when displaying I units on shelf, the retailer incurs total inventory cost of $H(I)$ \$ per year, where $H(I)$ is a general increasing and convex function. For a given contract (w, s) offered by the manufacturer, the retailer chooses her inventory level so as to maximize her own profit as follows

$$\underset{I \geq 0}{Max}\, \Pi^r(I) = (p - w)D(I) - H(I) + sI. \tag{7.14}$$

We can easily check that $\Pi^r(I)$ is concave and, hence, the retailer's optimal inventory level I^r can be found by solving the first order condition which yields,

$$(p - w)D'(I^r) = H'(I^r) - s. \tag{7.15}$$

When the channel is centrally controlled, the optimal inventory level I^c is determined by solving problem (7.14) with $w = c$ and $s = 0$. That is,

$$(p - c)D'(I^c) = H'(I^c).$$

In a decentralized setting, a (w, s) contract offered by the manufacturer will coordinate the channel if and only if it induces the retailer to choose the system-optimal inventory I^c. Such channel coordinating contracts can be characterized by (7.15). That is, for any given wholesale price w, the corresponding inventory-subsidy s is determined by

$$s = (p - w)D'(I^c) - H'(I^c).$$

Note that the manufacturer will have to possess information about the retailer's demand function (as well as her holding cost function) in order to offer a coordinating contract. In a sharp contrast, he does *not* need to know the demand function when the retailer's holding cost is linear, as shown in (7.10).

3. Two Competitive Retailers with Demand a Function of Aggregate Inventory

Here a manufacturer faces two competitive retailers (R1 and R2) who share a market with a constant selling price p. (We do not model the creation of this price in the duopolistic market since our focus is on the effect of stock levels on demand.) The total demand will depend on their aggregate inventory, i.e., $D(I_1 + I_2)$, where I_1 and I_2 are the inventory stocking levels of R1 and R2 respectively. The two retailers compete for this total demand based on their inventory levels. Say, R1 secures a portion $A(I_1, I_2)$ of $D(I_1 + I_2)$, where the fraction $A(I_1, I_2)$ is a function of I_1 and I_2. So, R2 has the rest, i.e., 1 - $A(I_1, I_2)$ of $D(I_1+I_2)$. $A(I_1, I_2)$ is increasing in I_1 and decreasing in I_2, capturing the shelf space-exposure competition between the retailers. This should be contrasted with splitting rules of the type analyzed by Lippman and McCardle (1997, Section 2), which are independent of the retailers' actions.

The manufacturer offers both retailers an *identical* price-plus-inventory-subsidy contract (w, s). Then, each retailer chooses her own inventory stocking level so as to maximize her own profit, knowing her profit also depends on the other retailer's action. Thus, their decisions affect each other's. That is,

$$\begin{aligned}\underset{I_1 \geq 0}{Max}\, \Pi_1(I_1, I_2) &= (p - w)A(I_1, I_2)D(I_1 + I_2) \\ &\quad -(h_1 - s)I_1, \ for\ R1; \qquad (7.16)\\ \underset{I_2 \geq 0}{Max}\, \Pi_2(I_1, I_2) &= (p - w)[1 - A(I_1, I_2)]D(I_1 + I_2) \\ &\quad -(h_2 - s)I_2, \ for\ R2, \qquad (7.17)\end{aligned}$$

where h_1 and h_2 are the inventory holding costs for Rl and R2, respectively.

Before analyzing the behavior of the two competitive retailers, we want to comment on a scenario where the two retailers (but not the manufacturer) are *centrally* controlled, so as to maximize the total profit $\Pi_1(I_1, I_2) + \Pi_2(I_1, I_2)$ by choosing I_1 and I_2 jointly. This centralized scenario is of interest on its own right. Also, the results obtained here will provide interesting insights into our later analysis of competitive/decentralized retailers. From (7.16) and (7.17), we have that

$$\Pi_1(I_1, I_2) + \Pi_2(I_1, I_2) = (p - w)D(I_1 + I_2) - (h_1 - s)I_1 - (h_2 - s)I_2.$$

Note that I_1 and I_2 appear here in the first term (i.e., total revenue net of wholesale costs) in an aggregate form, i.e., $I_1 + I_2$, but they appear separately in the holding cost terms. Thus, without loss of generality, if

R1 has a lower holding cost (i.e., $h_1 < h_2$), the best centralized decision will be to stock *only* at R1, and the optimal quantity will thus be the same as when R1 alone owns the whole market. Furthermore, we have that if the two retailers are identical (i.e., $h_1 = h_2$) and are centrally controlled, their total profit $\Pi_1(I_1, I_2) + \Pi_2(I_1, I_2)$ will depend only on their total inventory level $I_1 + I_2$ (i-e., it does not matter how any given total inventory is allocated between the retailers). The optimal total inventory will be the same as when one of the retailers is the only one operating on the given marketplace.

Let us now return to the competitive retailers setting. For this supply chain, we are interested in the following questions: For any given manufacturer contract (w, s), what are the retailers' equilibrium inventory decisions and their properties? Is the equilibrium unique? Can this two-parameter contractual arrangement coordinate the supply chain?

The general problem in (7.16) and (7.17) is too complex to analyze. To gain some concrete insights, we will consider the specific demand function of example 7.1:

$$D(I_1 + I_2) = a(I_1 + I_2)^b, a > 0, 1 > b > 0.$$

Such form of inventory-level-dependence was previously used (in a single retailer case) by Wang (1992) and Parlar et al. (1994). Note that $[dD(I)/dI]/[D(I)/I]$, the demand's inventory-elasticity, equals b. Second, we will use the proportional demand allocation model, which was motivated in the Introduction. That is,

$$A(I_1, I_2) = I_1/(I_1 + I_2), \text{ and so } 1 - A(I_1, I_2) = I_2/(I_1 + I_2).$$

Finally, we consider the case of two identical retailers; so we let $h_1 = h_2 = h$. With these specifications, problem (7.16) - (7.17) now reduces to

$$\begin{aligned} \underset{I_1 \geq 0}{Max}\, \Pi_1(I_1, I_2) &= (p - w) \cdot I_1/(I_1 + I_2) \cdot a(I_1 + I_2)^b \\ &\quad -(h - s)I_1, \ for\ R1; \qquad (7.18) \\ \underset{I_2 \geq 0}{Max}\, \Pi_2(I_1, I_2) &= (p - w) \cdot I_2/(I_1 + I_2) \cdot a(I_1 + I_2)^b \\ &\quad -(h - s)I_2, \ for\ R2. \qquad (7.19) \end{aligned}$$

The following theorem characterizes the competitive equilibrium for the game defined by (7.18)-(7.19):

Theorem 7.4 *The unique Nash equilibrium for each of the two retailers is to stock*

$$I^N = \left[a(1 + b)(p - w)2^{b-2}/(h - s)\right]^{1/(1-b)}. \qquad (7.20)$$

At the equilibrium, the system-wide inventory of both retailers will be

$$2I^N = [a(1+b)(p-w)/2(h-s)]^{1/(1-b)}. \tag{7.21}$$

Now, had one of the retailers occupied the entire market (and faced the same manufacturer contract), her optimal inventory would have been (specializing (7.9) to our particular demand function):

$$I^r = [ab(p-w)/(h-s)]^{1/(1-b)}. \tag{7.22}$$

Comparing this with (7.21), since $\infty > (1+b)/2b > 1$ for $0 < b < 1$, we have $2I^N > I^r$. That is,

Corollary 7.5 *For any given manufacturer contract (w, s), the total inventory displayed by two competitive retailers is always higher than that of a single retailer (or two centrally controlled retailers). Thus, competition generates* inefficiency *at the retail stage of the supply chain.*

Corollary 7.5 assumes that the prices in the monopolistic and duopolistic markets, which are not modelled, will be equal. In practice, the latter is likely to be lower due to competition. It is easy to check that in order for the Corollary's conclusion to be reversed it will have to be lower than $w + 2b(p-w)/(1+b)$, where p is the monopoly price.

The inventory displayed by one of the competitive retailers (I^N) can be higher or lower than that of a single retailer (I^r), depending on the value of the inventory elasticity parameter b. Note that the conclusions here also hold for a price-only contract, since that is merely a (w, s) contract specialized to $s = 0$.

Can a (w, s) contractual arrangement coordinate this manufacturer-competitive retailers supply chain? The answer is yes! To see this we only need to show that such a contract can bring the two retailers to choose the centralized or system-optimal inventory levels. But, from Proposition 7.6, we know that the *total* inventory of two retailers (independent of inventory allocation between them) maximizing the *system-wide* performance will be the same as that maximizing the performance of a system with a single retailer. Thus, to achieve channel coordination, one only needs to design (w, s) such that the decentralized total inventory given in (7.21) be equal to the centralized optimal inventory, which can be derived from (7.2) for a single retailer system (specializing to our specific demand function) as

$$I^c = [ab(p-c)/h]^{1/(1-b)}. \tag{7.23}$$

Thus, one can easily show,

Proposition 7.6 *If* (w, s) *is offered such that*

$$s = h - (1+b)(p-w)h/2b(p-c), \tag{7.24}$$

then $2I^N = I^c$.

In light of Corollary 7.5, this coordinating property of a (w, s) contract is particularly valuable: *it can actually eliminate the inefficiency generated by the presence of competing retailers within the supply chain!*

When coordination is achieved, each retailer's inventory is $I^N = I^c/2$. Substituting (7.24) into either (7.18) or (7.19), we can show that the retailers' profits will be

$$\Pi_1 = \Pi_2 = [a(I^c)^b/2 - (1+b)hI^c/4b(p-c)](p-w). \tag{7.25}$$

Thus, again, not only can a (w, s) contract coordinate the supply chain, but it can also achieve, by varying w, any desired allocation of the total channel profit between the manufacturer and the retailers.

In concluding this section, we compare the (w, s) coordination mechanisms for a single retailer to that for two competitive retailers. First, we see from (7.10) that in the single retailer case the manufacturer does not need to know anything about the demand function $D(\cdot)$ in order to design a coordinating contract, assuming profit allocation is not a concern. In a sharp contrast, to coordinate a supply chain with competing retailers, he *does* need to know the demand pattern, captured via the parameter b as shown in (7.24). Second, for a given wholesale price w, the two-retailer supply chain needs a smaller inventory subsidy s to be coordinated than the single retailer chain does. (With $(1+b)/2b > 1$, this can be seen by a direct comparison of (7.10) with (7.24).) This second point is not surprising since we know from Corollary 7.5 that two competing retailers will always hold more inventory with the same (w, s) contract. This observation seems to suggest that as the number of retailers grows the optimal subsidy declines. Future research needs to explore that behavior, and in particular whether the subsidy declines to zero as the number of retailers grows to infinity.

4. Two Competing Retailers with Demand a Function of Individual Inventory

In this section we model a situation where a customer chooses between R1 and R2 based on their relative displayed-inventory levels, but her demand quantity then depends solely on the inventory of the chosen retailer. Specifically, the competition process of the two retailers can be thought of as follows: When their displayed inventories are I_1

and I_2 units respectively, a portion $A(I_1, I_2)$ of the total N customers will choose R1, *each* with a demand quantity of $D(I_1)$, and the remaining portion 1-$A(I_1, I_2)$ of the N customers will choose R2, *each* with a demand quantity of $D(I_2)$.

With a manufacturer's (w, s)-type contract and inventory holding costs of h_1 and h_2 respectively, the retailers face the following decisions:

$$\begin{aligned} \underset{I_1 \geq 0}{Max}\, \Pi_1(I_1, I_2) &= (p - w) \cdot A(I_1, I_2) \cdot N \cdot D(I_1) \\ &\quad -(h_1 - s) \cdot I_1, \; for\ R1; \qquad (7.26) \\ \underset{I_2 \geq 0}{Max}\, \Pi_2(I_1, I_2) &= (p - w) \cdot [1 - A(I_1, I_2)] \cdot N \cdot D(I_2) \\ &\quad -(h_2 - s) \cdot I_2, \; for\ R2. \qquad (7.27) \end{aligned}$$

Consider the case where $A(I_1, I_2) = I_1/(I_1 + I_2)$, $N = a$, $D(I) = I^b$ with $0 < b < 1$, and $h_1 = h_2 = h$ (i.e., identical retailers). Then, (7.26)-(7.27) reduce to

$$\begin{aligned} \underset{I_1 \geq 0}{Max}\, \Pi_1(I_1, I_2) &= (p - w) \cdot I_1/(I_1 + I_2) \cdot a \cdot I_1^b \\ &\quad -(h - s) \cdot I_1, \; for\ R1; \qquad (7.28) \\ \underset{I_2 \geq 0}{Max}\, \Pi_2(I_1, I_2) &= (p - w) \cdot I_2/(I_1 + I_2) \cdot a \cdot I_2^b \\ &\quad -(h - s) \cdot I_2, \; for\ R2. \qquad (7.29) \end{aligned}$$

We note that here the two retailers essentially face the same market (i.e., the same a customers) as that in (7.18)-(7.19). The difference is that a customer who chooses, say, R1 will here contribute a demand of I_1^b, while in (7.18)-(7.19) it was $(I_1 + I_2)^b$. Thus the total demands are different in the two scenarios even when the inventory levels are the same.

It turns out that problem (7.28)-(7.29) is much more complex to analyze than problem (7.18)-(7.19). Instead of trying to fully characterize the response curves and equilibrium point(s), in the following we simply identify one specific equilibrium – the symmetric equilibrium, where the two retailers display the same amount of inventory. Intuitively, since the two retailers are identical, the most likely equilibrium, if any, should be symmetric.

From (7.28), a symmetric equilibrium, *if any,* can be found by substituting $I_1 = I_2 = I^N$ into

$$\begin{aligned} &\partial \Pi_1(I_1, I_2)/\partial I_1 = F_1(I_1, I_2) = \\ &a(p - w) I_1^b [b I_1 + (1 + b) I_2]/(I_1 + I_2)^2 - (h - s) = 0 \qquad (7.30) \end{aligned}$$

and solving for I^N. We thus obtain the *unique* solution

$$I^N = [a(1 + 2b)(p - w)/4(h - s)]^{1/(1-b)}. \qquad (7.31)$$

The next Theorem states that, if $b < 0.5$, $I_1 = I_2 = I^N$ is indeed an equilibrium point.

Theorem 7.7 (I^N, I^N) *constitutes a Nash equilibrium point of (7.28)-(7.29) if* $b < 0.5$.

Now, suppose that the two retailers are centrally controlled, so as to maximize their total profit by choosing jointly how much inventory each should stock (i.e., I_1 and I_2). We have the following result,

Proposition 7.8 *If the two retailers are centrally controlled, then, for any given total inventory, the best policy is to stock at only one of the retailers (and, hence, to close down the other).*

The intuition here is as follows. Since each customer's demand depends (increases) only on (in) the size of ONE pile, and as the total number of customers is constant in this model, then for any given total inventory, stocking all of it at one location will induce more demand than splitting it into two piles. While the demand function itself is concave, the profit function it gives rise to is convex (see proof in Appendix) and hence the boundary solution.

Can the manufacturer coordinate such a supply chain? It depends on how the retail stage operates. If the two retail locations are centrally controlled, Proposition 7.8 shows that one should close down one of the locations, and, thus, the retail stage acts just like the single-retailer. Then, as we have showed earlier, a (w, s) contractual arrangement offered by the manufacturer can coordinate the supply chain. But, when two retail locations co-exist through competition, Proposition 7.8 states that inefficiency/waste will occur within the retail stage. As a result, with a (w, s) contract arrangement, the manufacturer will not be able to coordinate the supply chain. Any attempt of coordination here must have, among other arrangements, the contingency of *physically* pooling inventory. In contrast, as we have shown in Section 3, supply chain coordination can be achieved with two competing retailers when customer demand depends on the "aggregate" inventory of both retailers. The fundamental difference is that there the allocation of inventory between the retailers does not in itself cause inefficiency.

5. Concluding Remarks

As argued by Moorthy (1993, p.182), "The ... interesting issues in channel competition arise from the effect of downstream (retail) competition on relations between the manufacturer and the retailers...". Our model indeed attempted to capture such interactions within a concrete setting. We did so (in the analysis of two competitive retailers) by viewing the system as that of a Stackleberg leader (the manufacturer) who considers the effect of its actions on the resulting Nash equilibrium of the competing retailers (for a similar modelling philosophy, see Gerchak and Wang 2000).

While our work was motivated and presented through demand's shelf space (or inventory displayed) dependence, the model is, in fact, rather general; I could correspond to any marketing effort. As such, our model can be viewed as a marketing problem as well as an operations problem. As pointed out, some components of our model – notably the demand split ratios – are often used by marketing researchers. The coordination issues and mechanism we addressed, however, seem new or different than models explored in the marketing literature.

Future research could deal with more general market-share models. For example, with n competing retailers, when each retailer i allocates shelf space I_i, retailer j has a market share of $\alpha_j I_j / \sum_i \alpha_i I_i$, where the coefficient α_i represent retailers i's relative *effectiveness* of shelf space utilization in attracting demand (e.g., Cooper 1993). Another would be $I_j^{\alpha_j} / \sum_i I_i^{\alpha_i}$.

Our models took retail price as given. A more general setting will have a demand which depends on price as well as inventory, and where price is a decision variable. In the duopolistic setting, that may call for a Bertrand-type approach. Since the prices will then depend on the type of market, the relations among the optimal inventory levels will be affected.

Our current models assumed a setting with complete information (though in the single-retailer setting the manufacturer did not always need to know the demand function and holding costs experienced by the retailer). A natural extension is to consider various scenarios where either the retailers or the manufacturer have some private information regarding costs or demand parameters. For recent work, see Ha (1998), Corbett and Tang (1999), Cachon and Lariviere (1999) and references therein.

Another extension to the current models is to consider stochastic demand which is influenced by inventory/shelf space. In a periodic review setting, Gerchak and Wang (1994) studied such models for centralized systems. Interesting issues of channel coordination might arise if one considers decentralized decisions with competition in such settings.

Acknowledgments

The authors wish to thank two anonymous referees for their valuable comments. Part of this work was done when the second author was at the University of Waterloo, and was supported by the Natural Sciences and Engineering Research Council of Canada. Some of the results appeared in a condensed form in the short paper by the same authors: "Supply Chain Coordination When Demand is Shelf-Space-Dependent," *Manufacturing & Service Operations Management,* 3(1), 82-87, 2001.

Appendix

Proof of Theorem 7.4: We note first that Anupindi et al. (1999, Theorem 4.1) provide two sufficient conditions for the existence and uniqueness of symmetric equilibrium. Unfortunately, we found that our model can not satisfy their second condition when $b < 0.5$. Now, to find the Nash Equilibrium, we will first characterize R1's reaction function denoted by $F_1(I_1, I_2)$; R2 will have an identical reaction function. Prom (7.18), it follows after some algebra that

$$\frac{\partial \Pi_1(I_1, I_2)}{\partial I_1} = F_1(I_1, I_2) = a(p-w)\frac{bI_1 + I_2}{(I_1 + I_2)^{2-b}} - (h-s) = 0, \qquad (7.A.1)$$

and

$$\frac{\partial^2 \Pi_1(I_1, I_2)}{\partial I_1^2} = -a(1-b)(p-w)\frac{bI_1 + 2I_2}{(I_1 + I_2)^{3-b}} < 0.$$

So, $\Pi_1(I_1, I_2)$ is concave in I_1 and, thus, (7.A.1) indeed defines the reaction function.

The following lemma partially characterizes the shape of $F_1(I_1, I_2)$; see Figure 7.A.1.

Lemma A.1: *In the* (I_1, I_2) *plane:*

1) $F_1(I_1, I_2)$ *passes through the following four points:*

P1: $I_1 = [ab(p-w)/(h-s)]^{1/(1-b)}$; $I_2 = 0$

P2: $I_1 = \ ab(p-w)(2-b)^{b-2}/(h-s) \ ^{1/(1-b)}$;
$I_2 = (1-b)I_1 = (1-b) \ ab((p-w)(2-b)^{b-2}/(h-s) \ ^{1/(1-b)}$

P3: $I_1 = I_2 = \ a(1+b)(p-w)2^{b-2}/(h-s) \ ^{1/(1-b)}$

P4: $I_1 = 0$; $I_2 = [a(p-w)/(h-s)]^{1/(1-b)}$

2) *The horizontal coordinate* I_2 *of P4 is longer than the vertical coordinate* I_1 *of P1.*

3) $F_1(I_1, I_2)$ *is increasing from P1 to P2, and is decreasing from P2 to P4.*

Proof of Lemma A.1:
Part 1) For P1, substituting $I_2 = 0$ into (7.A.1), we find I_1. P2 is the intersection of $F_1(I_1, I_2)$ with the line $I_1 = I_2/(1-b)$; so, together with (7.A.1),we can find I_1 and

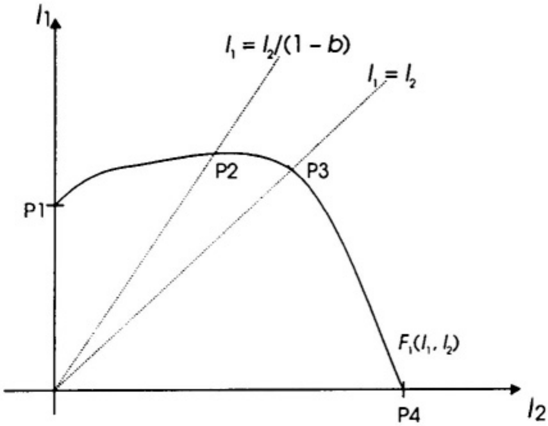


Figure 7.A.1. The Reaction Curve of Retailer 1

I_2. For P3 and P4, using (7.A.1) together with $I_1 = I_2$ and $I_2 = 0$ respectively, we can verify their coordinates as well.
Part 2) Since $0 < b < 1$, the result can be verified immediately.
Part 3) By implicit differentiation, one can show from (7.A.1) that

$$\frac{dI_1(I_2)}{dI_2} = \frac{(1-b)I_1 - I_2}{bI_1 + 2I_2}.$$

At P2, we have $I_1 = I_2/(1-b)$. Now, from P1 to P2, we have $I_1 > I_2/(1-b)$ and, hence, $dI_1(I_2)/dI_2 > 0$, which indicates that $F_1(I_1, I_2)$ is increasing. But, from P2 to P4, we know that $I_1 < I_2/(1-b)$, and so $dI_1(I_2)/dI_2 > 0$, and thus $F_1(I_1, I_2)$ must be decreasing.
End of Proof for Lemma 1.

With our characterization of the reaction function, we are ready to identify the Nash equilibrium point. If we place the reaction function $F_2(I_1, I_2)$ of R2 on the same (I_1, I_2) plane with $F_1(I_1, I_2)$, they will be symmetric across the line $I_1 = I_2$, since the two retailers are identical; see Figure 7.A.2. Thus, we immediately identify point P3, where $F_2(I_1, I_2)$ and $F_1(I_1, I_2)$ intersect, as one Nash equilibrium. To complete the proof of Theorem 7.4, we next show that P3 is the unique equilibrium point.

For any given value of I_2, let (I_1^1, I_2) and (I_1^2, I_2) be the corresponding points on $F_1(I_1, I_2)$ and $F_2(I_1, I_2)$ respectively; see Figure 7.A.2. We need to show that $I_1^1 \neq I_1^2$ except at point P3, that is, $F_1(I_1, I_2)$ and $F_2(I_1, I_2)$ do not intersect other than at P3.
Now, (I_1^1, I_2) satisfies (7.A.1), so we have

$$\frac{bI_1^1 + I_2}{(I_1^1 + I_2)^{2-b}} = \frac{h-s}{a(p-w)}. \tag{7.A.2}$$

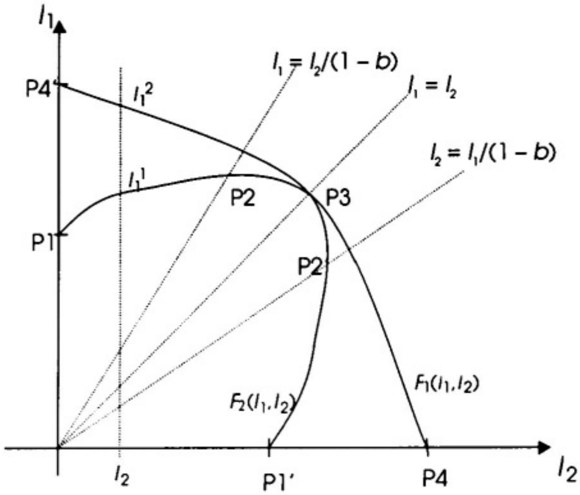


Figure 7.A.2. The Unique Nash Equilibrium Point

Similarly, by deriving $F_2(I_1, I_2)$ from (7.19) and then substituting (I_1^2, I_2), we have

$$\frac{I_1^2 + bI_2}{(I_1^2 + I_2)^{2-b}} = \frac{h - s}{a(p - w)}. \tag{7.A.3}$$

We next show that, except for P3, the segment P4'-P3 of $F_2(I_1, I_2)$ does not intersect with segment P1-P3 of $F_1(I_1, I_2)$, i.e., $I_1^1 \neq I_1^2$. (That P3-P4 of $F_1(I_1, I_2)$ does not intersect with P3-P1' of $F_2(I_1, I_2)$ will then follow immediately from the symmetry of $F_1(I_1, I_2)$ and $F_2(I_1, I_2)$.) Note that on P1-P3 of $F_1(I_1, I_2)$, we have

$$I_1^1 > I_2. \tag{7.A.4}$$

Now, suppose that $I_1^1 = I_1^2$. Then, from (7.A.2) and (7.A.3), we must have

$$bI_1^1 + I_2 = I_1^2 + bI_2 \Rightarrow I_2(1 - b) = I_1^2 - bI_1^1 = I_1^1(1 - b) \Rightarrow I_2 = I_1^1,$$

which contradicts (7.A.4)!

Proof of Theorem 7.7: We need to show that if R2 chooses $I_2 = I^N$, the best choice for R1 is $I_1 = I^N$ as well, that is, $\Pi_1(I^N, I^N) \geq \Pi_1(I_1, I^N)$ for all $I_1 \geq 0$. To that end, we will show that $\Pi_1(I_1, I^N)$, starting at $\Pi_1(0, I^N) = 0$, has an "S" shape, and it reaches its maximum at $I_1 = I^N$; see Figure 7.A.3.

The "S" shape of $\Pi_1(I_1, I^N)$ can be shown by studying its derivative function $\partial\Pi_1(I_1, I^N)/\partial I_1 = F_1(I_1, I^N)$. From (7.30), we have,

$$F_1'(I_1, I^N) = \frac{\partial F_1(I_1, I^N)}{\partial I_1}$$
$$= a(p - w)\frac{-b(1-b)I_1^2 - 2(1-b^2)I_1 I^N + b(1+b)(I^N)^2}{(I_1 + I^N)^3 I_1^{1-b}}. \tag{7.A.5}$$

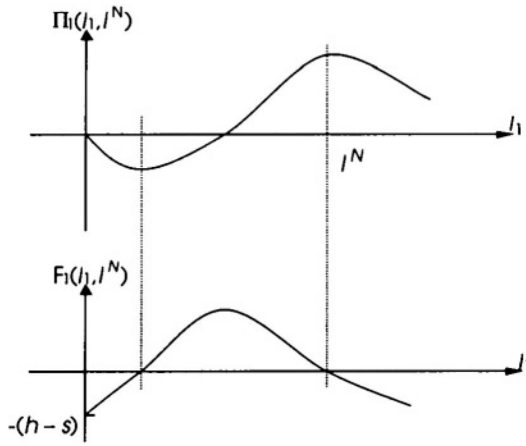


Figure 7.A.3. The Symmetric Equilibrium (I^N, I^N)

Now, since its numerator, namely, $-b(1-b)I_1^2 - 2(1-b^2)I_1 I^N + b(1+b)(I^N)^2$, is strictly decreasing, $F_1'(I_1, I^N)$ will start positive and then become and remain negative, which implies that $F_1(I_1, I^N)$ itself will initially be increasing and then become decreasing (i.e., it is unimodal). Furthermore, we can check from (7.A.5) that

$$F_1'(I^N, I^N) = -2(1-2b^2) < 0, \ if \ b < \sqrt{2}/2,$$

which implies that $I_1 = I^N$ is at the decreasing portion of $F_1(I_1, I^N)$. This, combined with $F_1(I^N, I^N) = 0$, indicates that $F_1(I_1, I^N)$, starting at $F_1(0, I^N) - (h-s) < 0$, increases to a positive value at some point before $I_1 = I^N$ and then decreases to zero at $I_1 = I^N$, and stay negative after $I_1 = I^N$. Thus, we have showed that $\Pi_1(I_1, I^N)$ has the "S" shape.

Now, to have $\Pi_1(I^N, I^N) \geq \Pi_1(I_1, I^N)$ for all $I_1 \geq 0$ and $b < 0.5$, we only need $\Pi_1(I^N, I^N) > 0$ for $b < 0.5$, which can be demonstrated simply by substituting $I_1 = I_2 = I^N$ into (7.28). This completes the proof.

Proof of Proposition 7.8: From (7.28) and (7.29), we can show that the total profit of the two retailers is

$$\Pi_1(I_1, I_2) + \Pi_2(I_1, I_2) = a(p-w)\frac{1}{I_1 + I_2}(I_1^{1+b} + I_2^{1+b}) - (h-s)(I_1 + I_2).$$

For any given total inventory, say k units, substituting $I_1 + I_2 = k$ and $I_2 = k - I_1$ into the above equation, we have

$$\Pi_1(I_1, I_2) + \Pi_2(I_1, I_2) = a(p-w)\frac{1}{k}[I_1^{1+b} + (k - I_1)^{1+b}] - (h-s)k.$$

Thus, it is easy to show that in order to maximize the total profit, the optimal solution is to set I_1 to be either zero (i.e., to stock only at R2) or k (i.e., to stock only at R1).

References

Anupindi, R. and Y. Bassok (1999) "Centralization of Stocks: Retailer vs. Manufacturer," *Management Science,* 45(2), 178-191.

Anupindi, R., Y. Bassok and E. Zemel (1999) "Study of Decentralized Distribution Systems: Part II - Application," *Working paper,* Kellogg Graduate School of Management, Northwestern University, Evanston, IL.

Baker, R.C. and T.L. Urban (1988) "A deterministic Inventory System with an Invenoty-Level-Dependent Demand Rate," *Journal of the Operational Research Society,* 39, 823-831.

Balakrishnan, A., M.S. Pangburn and E. Stavrulaki (2000) "Stack Them High, Let 'em Fly: Lot-Sizing Policies When Inventories Stimulate Demand," *Working paper,* Smeal College of Business Administration, Pennsylvania State University, University Park, PA 16802.

Cachon, G.P. (1999) "Competitive Supply Chain Inventory Management," Chapter 5 in Tayur et al. (1999).

Cachon, G.P. and M.A. Lariviere (1999) "Contracting to Assure Supply: How to Share Demand Forecasts in a Supply Chain," *Working Paper,* The Fuqua School of Business, Duke University.

Cooper, L.G. (1993) "Marketing Share Models," Chapter 6 in Eliashberg and Lilien (1993).

Corbett, C.J. and C.S. Tang (1999) "Designing Supply Contracts: Contract Type and Information Asymmetry," Chapter 9 in Tayur et al. (1999).

Corstjens, M. and P. Doyle (1981) "A Model for Optimizing Retail Space Allocation," *Management Science,* 27, 822-833.

Dana Jr., J.D. (2000) "Competition in Price and Availability when Availability is Unobservable," *Working paper,* Dept. of Management and Strategy, J.L . Kellogg Graduate School of Management, Northwestern University, Evanston, IL 60208.

Dana Jr., J.D. and K.E. Spier (2000) "Revenue Sharing, Demand Uncertainty, and Vertical Control of Competing Firms," *Working paper,* Dept. of Management and Strategy, J.L . Kellogg Graduate School of Management, Northwestern University, Evanston, IL 60208.

Deneckere, R. and J. Peck (1995) "Competition over Price and Service Rate When Demand is Stochastic: A Strategic Analysis," *RAND Journal of Economics,* 26(1), 148-162.

Eliashberg, J. And G.L. Lilien (eds.) (1993) *Marketing,* Vol. 5 of Handbooks in OR/MS, North Holland.

Eliashberg, J. and R. Steinberg (1993) "Marketing-Production Joint Decision-Making," Chapter 18 in Eliashberg and Lilien (1993).

Gerchak, Y. and Y. Wang (2000) "Revenue-Sharing vs. Wholesale-Price Contracts in Assembly Systems with Random Demand," *Working paper,* Department of Management Sciences, University of Waterloo.

Gerchak, Y. and Y. Wang (1994) "Periodic-Review Inventory Models with Inventory-Level-Dependent Demand," *Naval Research Logistics,* 41, 99-116.

Ha, A.Y. (1998) "Supply Contract for a Short-Life-Cycle Product with Demand Uncertainty and Asymmetric Cost Information," *Working Paper,* Yale School of Management.

Iglehart, D.L. and S.C. Jaquette (1969) "Multi-Class Inventory Models with Demand a Function of the Inventory Level," *Naval Research Logistics Quarterly,* 16, 495-502.

Jeuland, A. and S. Shugan (1983) "Managing Channel Profits," *Marketing Science,* 2, 239-272.

Johnson, E.L. (1968) "On (s, S) Policies," *Management Science,* 15, 80-101.

Kollat, D.T. and R.P. Willet (1969) "Is Impulse Purchasing Really a Useful Concept for Marketing Decisions?," *Journal of Marketing,* Jan., 79-83.

Kotler, P. (1984) *Marketing Management: Analysis, Planning and Control,* 5th ed., Printice-Hall.

Krugman, H.E. (1965) "The Impact of Television Advertising: Learning without Improvement," *Public Opinion Quarterly,* Fall, 349-356.

Lariviere, M.A. (1999) "Supply Chain Contracting and Coordination with Stochastic Demand," Chapter 8 in Tayur et al. (1999).

Lariviere, M. and E.L. Porteus (1999) "Selling to the Newsvendor," *Working Paper,* Fuqua School of Business, Duke University.

Lippman, S.A. and K.F. McCardle (1997) "The Competitive Newsboy," *Operations Research,* 45, 54-65.

Monahan, G.E. (1987) "The Structure of Equilibria in Market Share Attraction Models," *Management Science,* 33(2), 228-243.

Moorthy, K.S. (1993) "Competitive Marketing Strategies: Game-Theoretic Models," Chapter 4 in Eliashberg and Lilien (1993).

Moses, M. and S. Seshadri (1996) "Policy Mechanisms for Supply Chain Coordination," *Working paper,* Stern School of Business, New York University, New York, NY 10012.

Pasternack, B. (1985) "Optimal Pricing and Return Policies for Perishable Commodities," *Marketing Science,* 4, 166-176.

Pasternack, B. (1990), "The Newsboy Problem with Balking," Presented at ORSA/TIMS conference, Philadelphia, Oct. 1990.

Parlar, M., S. Bar-Lev and D. Perry (1994) “On the EOQ Model with Inventory-Level-Dependent Demand Rate and Random Yield,” *Operations Research Letters,* 16, 167-176.

Schary, P.B. and B.W. Becker (1972) “Distribution and Final Demand: The Influence of Availability,” *Mississippi Valley Journal of Business and Economics,* 8, 17-26.

Spengler, J. (1950) “Vertical Integration and Anti-Trust Policy,” *Journal of Political Economy,* 58, 347-352.

Tayur, S., R. Ganeshan and M. Magazine (1999) *Quantitative Models for Supply Chain Management,* Kluwer Academic Publishers.

Tsay, A.A. and N. Agrawal (1998) “Supply Chain Dynamics Under Price and Service Competition,” Department of Operations & Management Information Systems, Santa Clara University

Wang, Y. (1992) *Inventory Models with Inventory-Level-Dependent Demand,* M.A.Sc Thesis, Department of Management Sciences, University of Waterloo.

Chapter 8

SAM: A DECISION SUPPORT SYSTEM FOR RETAIL SUPPLY CHAIN PLANNING FOR PRIVATE-LABEL MERCHANDISE WITH MULTIPLE VENDORS

Stephen A. Smith
Department of Operations and Management Information Systems
Leavey School of Business, Santa Clara University
Santa Clara, CA 95053
ssmith@scu.edu

Narendra Agrawal
Department of Operations and Management Information Systems
Leavey School of Business, Santa Clara University
Santa Clara, CA 95053
nagrawal@scu.edu

Andy A. Tsay
Department of Operations and Management Information Systems
Leavey School of Business, Santa Clara University
Santa Clara, CA 95053
atsay@scu.edu

Abstract A number of retail firms use a "private-label" strategy in which merchandise is sold under a brand name exclusive to the retail firm, but manufactured by one or more independent vendors. While offering a number of benefits, this approach also poses a different set of supply chain challenges than manufacturer-brand-based retailing, in that the retail firm must take a more active role in organizing and coordinating the planning and materials management activities in a supply base that is often dispersed and heterogenous.

J. Geunes et al. (eds.),
Supply Chain Management: Models, Applications, and Research Directions, 163–198.

This chapter describes a methodology for planning capacity commitments, scheduling shipments, and managing inventory for an assortment of private-label retail merchandise produced by multiple vendors. The vendors differ in their lead time requirements, costs, and production flexibility. Product demand is uncertain, and fluctuates over time. We develop an optimization model to choose the production commitments that maximize the retailer's expected gross profit, given market demand forecasts and vendors' capacity and flexibility constraints. The model has been incorporated into a PC-based decision support system called the *Sourcing Allocation Manager* (SAM). This was developed in collaboration with supply chain planners at a global retailer of seasonal and fashion merchandise, and has been tested by buyers at two major retailers.

Keywords: Sourcing Strategy, Retailing, Capacity Planning, Multi-item Inventory Planning

1. Introduction

A number of retail firms use a "private-label" strategy in which merchandise is sold under a brand name exclusive to the retail firm, but manufactured by one or more independent vendors. This practice can allow a retailer to avoid the premium charged by brand-name vendors, fill gaps in its product assortment, exercise greater control over product attributes, gain leverage in the manufacturer-retailer balance of power, and convert product brand loyalty to store loyalty. For well-received products, there are additional benefits to be enjoyed from being the exclusive seller. However, this also poses a different set of supply chain challenges than manufacturer-brand-based retailing, in that the retail firm must take a more active role in organizing and coordinating the planning and materials management activities in a supply base that is often dispersed and heterogenous[1]. As a result, some such retail firms have become increasingly interested in tools and techniques for effective supply chain management and design. This is the case with the retailer (a multinational firm with several billion dollars of annual revenue from private-label sales) that approached us with the business problem motivating the research described here.

[1] Private-labeling poses a number of marketing challenges as well. The retailer takes sole responsibility for brand management tasks such as advertising and creating store displays, and foregoes manufacturer-sponsored provisions that mitigate market risks, such as return privileges and price protection. Our intent is not to address the question of when a retailer should use private-label, but to provide guidance on supply chain planning when this strategy is pursued.

We consider the problem of how to optimally plan and execute the sourcing of seasonal and fashion private-label merchandise carried by department stores and specialty retailers. For a given selling season, the sourcing decisions, typically made by the retail buyer responsible for each merchandise department, include the following components: (1) purchases of raw materials (e.g., fabric) for use by vendors, (2) supply contracts and production commitments with vendors, (3) a weekly plan for sales, shipments, and inventory, and (4) adjustments based on subsequent market information. This research develops a formal planning methodology for this decision problem that accommodates multiple products and multiple suppliers, and explicitly accounts for demand uncertainty and adjustments to the plan during the season. The resulting optimization model has been embedded within a PC-based decision support system named the *Sourcing Allocation Manager* (SAM).

A more theory-oriented treatment of this modeling research is presented in Agrawal et al. (2001). Parts of that document describing the model formulation are included here for the reader's convenience, but those who are interested in such a perspective and an extensive numerical case study should refer to that paper. This chapter focuses on the software implementation and how the business environment influenced the design of the graphical user interface.

The Business Setting

Many of the challenges of this application are due to attributes of the demand patterns and the supply base, and how they interact. Demand in this environment typically fluctuates sharply throughout the year. This is exemplified by the data in Figure 8.1, which illustrates recent sales for a men's casual slacks product.

This type of demand becomes most challenging when production capacity is constrained, which is commonly the case in this industry. Specifically, demand during the peak Fall ("Back to School") and Christmas seasons typically exceeds available manufacturing capacity, while surplus capacity tends to exist during the Spring and Summer. Producing in advance of peak periods improves the ability to meet demand, but creates inventory buildup arid requires that commitments to production and fabric purchases be made under greater uncertainty.

Sourcing strategies must also reflect the performance capabilities of the supply base. In most cases there are a variety of possible vendors that differ in costs, lead times, and flexibility of production. Vendors with the lowest cost generally offer virtually no flexibility with respect to capacity commitments. These vendors tend to have long lead times for booking capacity (e.g., nine months), shipment times of several weeks,

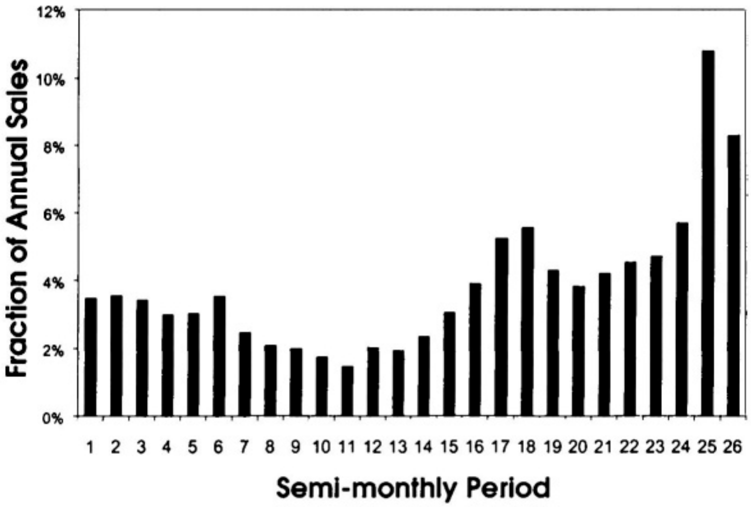


Figure 8.1. Seasonal Patterns in Demand[2]

and often require that the total production be allocated relatively evenly throughout the year. More responsive vendors may have shorter lead times and allow greater flexibility vis-a-vis production commitments. Additionally, different vendors may be willing to store limited amounts of finished product prior to delivery for a fee.

Retailers tend to leverage a portfolio of such vendors, resulting in supply chains such as that shown in Figure 8.2. The portfolio approach enables strategies such as exploiting lower cost production for the more predictable segment of demand, while sourcing the more speculative segment via the more flexible, but more costly, vendors. Operationalizing this strategy in a multi-product, multi-vendor setting is nontrivial, and is further complicated by many production and logistical constraints described later. This was our retail collaborator's motivation in sponsoring this project. In fact, our methodology is unique in its focus on designing contracts with a portfolio of vendors that simultaneously exploits the comparative advantages of each, as opposed to selecting a single most desirable vendor.

Research Contribution

Relative to previous academic research detailed in Section 2, our for-

[2]Since the retailer providing this data aspires to and usually achieves very high fill rates for this product, the difference between sales and demand is insignificant.

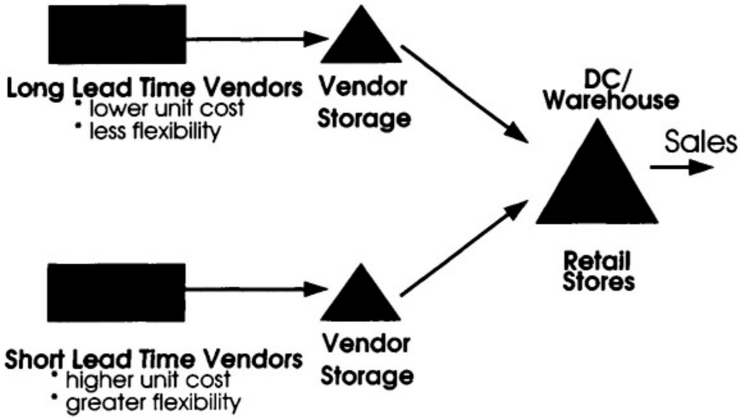


Figure 8.2. Supply Chain Structure

mulation of the multi-vendor sourcing problem is novel in representing the complex constraints and changing states of information under which different sourcing commitments must be made. We address numerous issues associated with the design of the supply chain, and provide insight into a universal question in sourcing: how to balance unit costs versus supplier attributes such as flexibility. Overall, our model builds on the key aspects of the literature described in Section 2, incorporating seasonal patterns in demand and detailed production and logistical constraints in a stochastic demand environment with forecast updating. While subsets of these issues have been treated previously, we believe our formulation to be unique in addressing all of them simultaneously.

Our formulation evolved in close collaboration with retail practitioners, whose involvement occurred at two different levels. A committee of senior executives from different firms regularly reviewed our assumptions and problem framing to ensure the broad applicability of our model to a variety of retail settings. However, the specifics were developed in collaboration with executives and buyers at a particular retailer, who confirmed that our level of detail captures the key complexities faced by retail planners. Their help was especially useful in identifying the cost tradeoffs and constraints most important for sensitivity analysis, leading to variable and constraint modifications that allowed discovery and presentation of the most critical shadow prices. Furthermore, feedback from these buyers and planners was instrumental in the incorporation of our model into a decision support software package with a graphical user interface. Given the depth and breadth of the practitioners' participation, we believe this model to be widely applicable to retail firms that manage the sourcing and production of private-label merchandise, and to certain nonretail firms as well.

Organization of This Chapter

The remainder of this chapter is organized as follows. Section 2 reviews the relevant literature. Section 3 details the mathematical formulation of the optimization model, discussing in depth the assumptions we made to capture the salient features of the particular retail environment. Section 4 describes the decision support software, the business issues motivating the design features of the user interface, and summarizes our retail collaborator's experiences with the software, and Section 5 concludes.

2. Literature Review

The use of formal decision models in aggregate production planning has a long tradition, and has been the subject of hundreds of academic studies. See Silver and Peterson (1985) for a textbook treatment and some historical background. A review of the academic literature is provided in Nam and Logendran (1992), and a survey of the usage of such models in practice is provided in Buxey (1993) and Buxey (1995). The predominant optimization approach is based on linear programming (LP), which allows for non-stationary but deterministic demand, and can handle large numbers of products simultaneously. Forecast uncertainty and information updating are usually dealt with only in an indirect fashion, by using a rolling-horizon implementation of a snapshot deterministic solution (the formal term for this is "Open-Loop Feedback Control", cf. Bertsekas (1976)), and also perhaps through the specification of safety stock levels, usually exogenously (e.g., Guerrero et al. (1986), Gunther (1982), Heath and Jackson (1994), Miller (1979)).

More direct treatment of demand uncertainty is called for in the retail setting, especially where hard-to-forecast fashion or style goods are involved. This can be provided by newsvendor-style models, but at the expense of sacrificing the dimensionality and detailed constraint structure that can be supported by LP formulations. In this type of approach, the entire selling season for a product is summarized as a small number (possibly one or two) of random variables with known joint probabilities. This allows analytic incorporation of forecast uncertainty into production planning (e.g., Crowston et al. (1973), Hartung (1973), Hausman and Peterson (1972), Murray and Silver (1966), Ravindran (1972), Wadsworth (1959), and more recently Brown and Lee, Donohue (2000), Eppen and Iyer (1997), Fisher and Raman (1996)), albeit in stylized ways. Various approaches to obtaining the probability distributions of these demand random variables, especially for fashion products, are proposed by Chang and Fyffe (1971), Hausman and Sides (1973), Hertz and Schaffir (1960), Riggs (1984), Riter (1967), and Wolfe (1968).

The efforts closest in spirit to our work are Bitran et al. (1986), Eppen et al. (1989), Kira et al. (1997), and Nuttle et al. (1991). The first three are based on mathematical programming, while the fourth takes a simulation approach. We discuss them briefly below.

In Bitran et al. (1986), the authors perform multi-period production planning for families of consumer electronics products which in turn consist of specific items. Setup costs for switchovers between families are such that each family will be run only once during the season, while switchovers between items within a family are assumed costless. Demand occurs in the last period, and estimates of this demand are revised each period. Demands for all items are assumed to be normally distributed, and the standard deviation of forecast error at each time period is known, given by an arbitrary, decreasing sequence of numbers which must be provided as data. The updated forecasts at each period also follow a joint normal distribution, with a known covariance matrix. The exact problem is a difficult-to-solve, stochastic mixed-integer program, for which the authors develop a deterministic mixed-integer approximation. While both their model and ours consider multi-product planning with forecast updates, the respective areas of emphasis differ. Whereas they take production capacity as given and then determine how to schedule the production of a variety of items, we consider as decision variables the capacities to be reserved with a variety of vendors at different points in time. They model the operations within a single factory at a greater level of detail, whereas our scope spans multiple vendors' factories as well as the retailer's distribution center, and includes the scheduling of shipments from the former to the latter. Their representation of item demand is more general but also data-intensive. We pursue a discrete simplification of forecast dynamics as part of an overall strategy of retaining a basic LP structure that allows an exact solution in real time.

In Eppen et al. (1989) a model is developed for General Motors to aid in making decisions about capacity for several lines of automobiles produced in multiple factories. A general sequence of events is considered in each of five years: (1) the available capacity is configured in terms of tooling the production lines for specific products, (2) demand occurs, and (3) a production plan is implemented that attempts to meet the demand given the capacity configuration. Demand uncertainty is represented by defining three different "scenarios" for each year that specify the demand and price for individual products. Scenario probabilities are assigned, and are assumed to be independent from year to year. The resulting optimization problem is a mixed-integer program that maps out individual sample paths of all possible scenario combinations. This scenario approach is similar to our representation of demand uncertainty. However, our production decisions are based on imperfect demand signals, while theirs assume that all uncertainty has been resolved. Further, our notions of capacity are slightly different. Their optimal capacity configuration is selected from a discrete number of predefined possibilities, hence the integer variable structure. Ours is chosen from a simplex region defined by a variety of constraints that explicitly represent features of the business relationship between the retailer and each vendor.

In Kira et al. (1997), the authors use a probability structure similar to that in Eppen et al. (1989), with a single-factory production environment that is much simpler than ours. Capacity planning is not treated, and the nuances of managing a supply chain composed of multiple, independently-managed physical nodes are not incorporated into their formulation.

In Nuttle et al. (1991) a software application called "The Sourcing Simulator" is described, which was developed by researchers at North Carolina State University in concert with the Textile/Clothing Technology Corporation and the American Textile Partnership-Demand Activated Manufacturing Architecture (AM-TEX-DAMA) project. This treats the same industry setting as we do, and makes many similar assumptions in addressing the question of how the replenishment frequency and lead time of a vendor affects a retailer's performance. This purely descriptive simulation approach allows a detailed representation of certain aspects of the setting, especially in the range of allowable replenishment strategies and consumer behavior. However, because it assumes single-sourcing (with the single vendor abstracted as simply a lead time and reorder frequency), it cannot simultaneously allocate production across a portfolio of time-phased vendors. Like the three models described previously, the scope of this formulation is largely confined to a single firm. Nevertheless, various studies based on this model (Hunter et al. (1992), Hunter et al. (1996) and King and Hunter (1996)) have validated the importance of the ability to react to improved demand information, which is a key rationale for the sourcing strategies that we model.

3. Model Specification

This section outlines the mathematical formulation of the planning problem faced by a retailer leveraging a portfolio of time-phased vendors. Our discussion uses the language of apparel retailing because this is our sponsor firm's primary line of business. However, we believe our underlying methodology to be broadly applicable to other product settings.

3.1 Timeline of Events and Information Assumptions

In chronological order, the critical time points for the retailer's sequential decision problem for a specified "selling season[3]" are as follows:

t_0 = time at which initial vendor commitments and fabric purchases[4] are made

t_1 = second time at which commitments to vendors are made, for those vendors allowing capacity decisions to be deferred to this time[5]

t_b = beginning of selling season

t_f = end of selling season, when actual demand becomes known.

We assume that our model analysis is performed at some time at or before t_0 for a selling season that spans the horizon (t_b, t_f). The retail planner's information regarding demand evolves continuously over time, shaped by economic forecasts, new fashion and color trends, and observed sales results for similar products. However, for our formulation it is only necessary to define the possible states of information at the specific points in time at which decisions are made. Evaluation of the expected profit also requires knowledge of the actual demand information at time t_f. To represent the evolving demand information, we define the following random variables[6]:

---

[3]This might correspond, for example, to the Fall season (running from roughly August through January) or the Spring season (February through July). For certain merchandise, some retailers use four or more shorter seasons per year. In some instances a season may be as short as 8 weeks.

[4]In many cases the fabric is purchased by the retailer and shipped to vendors for cutting and sewing. This provides control of raw material quality and leverages the buying power that a major retailer enjoys.

[5]Our discussions with the retailer's production planning managers indicated that two decision points (times t_0 and t_1) are adequate for a typical apparel planning decision process. However, the formulation can easily be extended to include more decision points by simply adding more variables to the model.

[6]For example, we have assumed that the initial demand information for any product at time t_0 is deterministic, i.e., X_0 has only one possible value. At time t_1, the demand information demand has three possible values based on what has been observed since t_0: High, Medium, or Low, with different probabilities. The remaining uncertainty about the actual demand is

$X_k \equiv$ a random variable corresponding to the market demand information that the retailer has at time t_k, for $k = 0, 1, f$.

At each time point, X_k has a discrete set of possible values. Finally, at time t_f, the actual demand corresponds to one of a discrete set of demand scenarios. We define the following probability distributions to describe the likelihood of observing particular sequences of demand information:

$p(\xi_1) \equiv P\{X_1 = \xi_1\}$ for each possible ξ_1 value at time t_1

$p(\xi_f|\xi_1) \equiv P\{X_f = \xi_f|X_1 = \xi_1\}$ for each possible combination of ξ_1 and ξ_f, and

$p(\xi_1, \xi_f) \equiv p(\xi_f|\xi_1)\,p(\xi_1) =$ the joint probability of X_1 and X_f.

Clearly, this structure can be generalized to characterize information that is revealed in any number of stages, but we will describe only the two-stage case since that corresponds to our particular application.

Market "scenarios" are frequently used by retailers in developing marketing plans for alternative contingencies[7]. We extend this concept to include market demand information that is revealed in stages, resulting in the sequential stochastic decision model illustrated in Figure 8.3. The underlying assumption is that as the selling season gets closer, the sales estimates in the plan improve for several reasons. For example, there is new sales information for related products. Also, updated sales estimates are at least in part the result of revisions in the merchandise plan, e.g., deciding to feature more or less of a particular type of merchandise, giving it a more or less prominent display and floor space, etc. For

then described by the conditional probability distribution , and is not completely resolved until the end of the selling season at time t_f. We also note that this modeling structure is easily generalizable to include additional stages of information and decision points.

[7]We model demand uncertainty through discrete scenarios for three reasons. The first reason is analytical tractability. Modeling uncertainty using continuous random variables would rule out certain complexities categorically declared by our corporate collaborators to be essential attributes of their business setting. The second reason is consistency with common managerial practice. Our corporate collaborators indicated that their planning methodology often requires the articulation of "worst case," "most likely," and "best case" scenarios for market uncertainties. However, in the past these scenarios have typically been used only for financial planning, due to a lack of technical know-how for translating them into contingency plans for vendor and production management. The third reason is that there is an established precedent in the literature for using scenarios to model uncertainty in a variety of contexts. As described in Section 2, Eppen et al. (1989) and Kira et al. (1997) used a scenario approach similar to ours for capacity planning. Discrete demand scenarios were used in Smith et al. (1998) to obtain optimal inventory and promotional plans for retail chains. Of course, there is a rich tradition in the financial economics literature of modeling uncertainty in the prices of stocks and securities this way (cf. Cox and Rubinstein (1985)). More recently, Huchzermeier and Cohen (1996) have used discrete scenarios to study the operations management implications of exchange rate fluctuations.

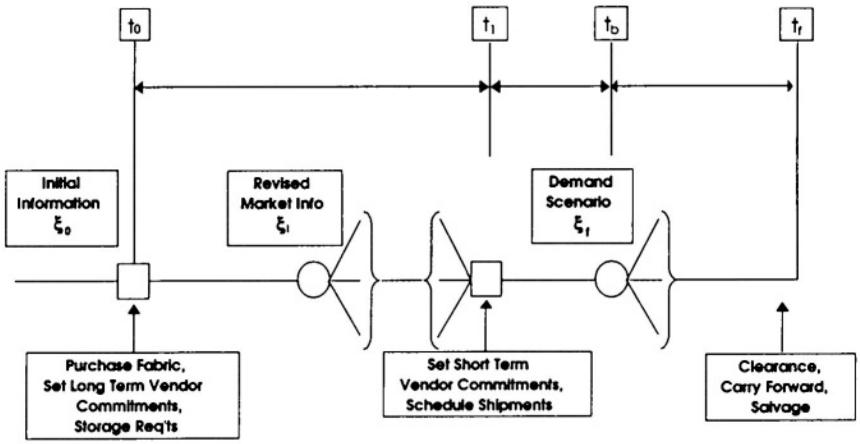


Figure 8.3. Decision Tree for Production Planning

changes of this type, there is a good base of experience for the buyers to update their subjective estimates of the demand. This determines the conditional probabilities $p(\xi_f|\xi_1)$. Note that in principle the same method can use early sales results to update the demand probabilities (after the selling season begins) and in fact, our formulation approach is compatible with Bayesian updating of the probabilities of the discrete demand levels based on early sales results. However, for this application the vendor deadlines did not permit changes in capacity commitments after the start of the selling season, other than changes in the color, style, or size mix. Since our model is meant to support capacity planning at an aggregate level, this is appropriate for our application[8]. With some assistance from the authors, the retail planners at two major retailers were able to subjectively estimate the required probabilities.

3.2 Decision Variable Definitions

The following indices will be used for variabledefinitions: j for products, i for vendors, t for the time increment used for production, ship-

[8]Most papers that consider updating of forecasts in a model of reasonably realistic detail only consider updating prior to the occurrence of any sales. This includes the mathematical-programming-based models most similar to ours, as described in Section 2. Those models that do accommodate forecast updating based on in-season sales tend to have very simplified inventory analysis that would not scale to the constraint and decision variable complexity in our decision model (e.g. Chang and Fyffe (1971), Crowston et al. (1973), Fisher and Raman (1996), Hartung (1973), Murray and Silver (1966)).

ment, and sales decisions (typically weeks), and q for the time increment used for reservation of capacity (typically quarters). In this model, we assume that the term "product" refers to an aggregation of styles, not to an individual SKU (distinguished by style/size/color). Variable names in upper case represent decision variables, while those in lower case or Greek symbols are fixed parameters.

The main basis for classifying vendors (into "short lead time" and "long lead time" types) is the time at which commitments for each product must be made. This is denoted by

$\tau_{ij} \equiv$ time at which a commitment is required by vendor i for product j,

and the corresponding state of retailer information

$X_{ij} \equiv$ demand information available at time τ_{ij}, which takes on discrete values ξ_{ij}.

For our implementation, $\tau_{ij} = t_0$ or t_1, since these are the only production commitment time points. It follows that X_{ij} is either X_0 or X_1 for every combination of i and j.

For each possible (ξ_1, ξ_f) combination, the production and inventory variables are defined as follows:

$F_j \equiv$ fabric commitment (in yards) made at time t_0 for product j

$P_{ij}(t|\xi_{ij}) \equiv$ production by vendor i of product j during period t

$Z_i(q|\xi_1) \equiv$ total production by vendor i during quarter q

$Z_j^F(\xi_1) \equiv$ yards of fabric actually used for product j

$M_{ij}(t|\xi_1) \equiv$ beginning inventory of product j stored by vendor i in period t

$S_{ij}(t|\xi_1) \equiv$ quantity of product j shipped from vendor i in period t

$U_j(t|\xi_1, \xi_f) \equiv$ retailer's unit sales of product j in period t

$I_j(t|\xi_1, \xi_f) \equiv$ retailer's beginning inventory of product j in period t

The decision variables depend on the information states in different ways, i.e., what information is known when each variable's value is specified. These dependencies determine the dimensionality of the variables. We denote this dependence explicitly in our formulation, using the "|" notation. For example, since the production schedule of an item i at a vendor j is fixed at time τ_{ij}, when the state of information is ξ_{ij}, the corresponding vendor production variables are denoted as $P_{ij}(t|\xi_{ij})$. The total production and total fabric usage depend upon ξ_1 because they are defined for both short and long lead time production decisions. Similarly, the vendors' inventory and shipment decisions depend upon ξ_1 because that is the information available to the vendor when the shipping decisions are made. However, the realized unit sales, and consequently the retailer's on hand inventory, depend on both ξ_1 and ξ_f. This is because the on-hand inventory depends on both the actual demand scenario and all the production decisions, some of which depend on ξ_1. Since the unit sales are affected by the inventory level, this depends on ξ_1 and ξ_f as well. In the LP optimization, the information states ξ_1 and ξ_f are simply treated as additional "subscripts" on variables.

3.3 Inventory Balance Equations and Production Constraints

The production, inventory, and shipping variables are related to each other by the following inventory balance equations for the retailer and vendors:

$$I_j(t+1|\xi_1,\xi_f) = I_j(t|\xi_1,\xi_f) + \sum_{i|t\geq\tau_{ij}+l_i} S_{ij}(t-l_i|\xi_1) - U_j(t|\xi_1,\xi_f),$$
$$\text{for all } i,j,\xi_1,\xi_f,t, \qquad (8.1)$$

where l_i is the shipping delay for vendor i,

$$M_{ij}(t+1|\xi_1) = M_{ij}(t|\xi_1) + P_{ij}(t|\xi_{ij}) - S_{ij}(t|\xi_1),$$
$$\text{for all } i,j,\xi_1,t. \qquad (8.2)$$

When the states of information "subscripts" in one constraint are different for different variables, the variable with fewer subscripts simply keeps the same value for a subset of the equations.

For simplicity, our model considers only the total inventory in the retailer's system, as opposed to inventory levels in individual stores[9]. This assumes that inventory is generally balanced across the stores, and

[9]Once the merchandise reaches the retailer's distribution center (DC), it is usually distributed to the stores and displayed for sale within two to three days. In order to maximize the productivity per square foot, there is generally little storage space in stores, and all store merchandise is placed on display for sale as quickly as possible. The only significant delays in this type of supply chain arise from production commitment lead times, which are usually several months, and shipping times, which may be several weeks for surface shipments.

is appropriate because inventory is re-balanced during the season by allocating replenishments to the stores that most need additional stock. For some merchandise, transshipments are made from one store to another to balance the inventory, but only if the repackaging and shipping costs can be justified.

Constraints on each vendor's storage space can be represented as

$$\sum_j \nu_j M_{ij}\left(t|\xi_1\right) \le w_i\left(t\right) \equiv \text{vendor } i\text{'s maximum storage for period } t,$$

$$\text{for all } \; i, t, \xi_1 \tag{8.3}$$

where $\nu_j \equiv$ storage space required per unit of product j.

A retailer may also specify an upper bound on the amount of inventory contained within its system[10]. This can be specified by

$$\sum_j \nu_j I_j\left(t|\xi_1, \xi_f\right) \le w^R(t) \equiv \text{retailer's maximum storage for period } t,$$

$$\text{for all } \; t, \xi_1, \xi_f \tag{8.4}$$

The initial and final inventories may also be required to satisfy constraints of the form:

$I_j\left(t_b|\xi_1, \xi_f\right) \ge i_j^0 \equiv$ minimum initial retailer inventory for product j for all ξ_1, ξ_f

$I_j\left(t_f|\xi_1, \xi_f\right) \ge i_j^f \equiv$ minimum final retailer inventory for product j for all ξ_1, ξ_f

$M_{ij}\left(t_b|\xi_1\right) \ge m_{ij}^0 \equiv$ minimum initial inventory of product j at vendor i for all ξ_1

$M_{ij}\left(t_f|\xi_1\right) \le m_{ij}^f \equiv$ maximum final initial inventory of product j at vendor i for all ξ_1.

The initial inventory i_j^0 must be sufficient to create an attractive display of merchandise with which to begin the selling season. For continuing, or "basic" products, the minimum final inventory i_j^f may be set to the desired initial inventory for the subsequent season. The vendor's initial inventory m_{ij}^0 can be used to satisfy demand in the current season, while the final inventory m_{ij}^f is available for the subsequent season.

[10]This can represent either a physical or budget restriction. In the latter case, ν_j will have a different meaning.

For some aspects of aggregate production planning, managers use quarters as the appropriate increment of time. Using $q(t)$ to denote the quarter corresponding to a time period t, the following relationship tallies the total production of vendor i within a quarter y:

$$\sum_{j} \sum_{t|q(t)=y} \kappa_j P_{ij}(t|\xi_{ij}) = Z_i(y|\xi_1), \quad \text{for all } i, y, \xi_{ij} \tag{8.5}$$

where $\kappa_j \equiv$ production capacity required per unit of product j. This enables us to model quarterly constraints. For instance, to ensure diversification a vendor may be willing to commit only a fraction of its quarterly capacity to a single retailer. On the other hand, less flexible vendors may also insist on a minimum quarterly production commitment from the retailer as a condition for doing business. These can be included as follows:

$$\underline{k}_i(q) \le Z_i(q|\xi_1) \le \overline{k}_i(q), \quad \text{for all } \ i, q, \xi_1 \tag{8.6}$$

where the bounds do not depend on the demand information. To achieve the economic benefits of level production, certain vendors also permit only limited changes of total production from quarter to quarter, which can be expressed as follows:

$$(1-\alpha_i) Z_i(q-1|\xi_1) \le Z_i(q|\xi_1) \le (1+\beta_i) Z_i(q-1|\xi_1), \quad \text{for all } i, q, \xi_1 \tag{8.7}$$

where $0 \le \alpha_i \le 1$ and $\beta_i \ge 0$. In general, vendors that allow later commitments also typically allow greater quarter-to-quarter flexibility (larger α_i and β_i parameters).

Production is also constrained by the fabric procurement decision as follows:

$$\sum_{y} \sum_{j} \sum_{t|q(t)=y} \kappa_j^F P_{ij}(t|\xi_{ij}) = Z_j^F(\xi_1) \le F_j, \quad \text{for all } \ j, \xi_{ij} \tag{8.8}$$

where $\kappa_j^F \equiv$yards of fabric required per unit of product j.

3.4 Modeling Product Demand

The demand pattern for each product over time is an input to the model that is conditional on the demand scenario ξ_f, denoted as follows:

$d_j(t|\xi_f) \equiv$ actual demand for product j in period t.

To specify these values, we used a forecasting model form that has been applied successfully to retail sales forecasting. Econometric marketing studies have found that multiplicatively separable models of the form

$$\begin{pmatrix}\text{Period } t \text{ demand}\\ \text{for product } j\end{pmatrix} = \begin{pmatrix}\text{Total season demand}\\ \text{for product } j\end{pmatrix} \cdot \begin{pmatrix}\text{Seasonality}\\ \text{effect at } t\end{pmatrix} \cdot \begin{pmatrix}\text{Marketing}\\ \text{effects at } t\end{pmatrix}$$

fit observed retail sales data well (Achabal et al. (1990), Kalyanam (1996)). Thus we let

$$d_j(t|\xi_f) = b_j(\xi_f) \cdot f_j(t) \cdot \rho_j(t) \tag{8.9}$$

where

$b_j(\xi_f) \equiv$ full-season demand for product j under demand scenario ξ_f

$f_j(t) \equiv$ fraction of total demand for product j that occurs in period t

$\rho_j(t) \equiv$ marketing effects for product j during period t, including price/ advertising effects.

This approach greatly reduces the model dimensionality by confining the effect of information updating to the full-season demand, which is a scalar. The full set of relative seasonality factors $f_j(t)$, such as that shown in Figure 8.1, generally do not require updating. Similar representations of demand have been used by Chang and Fyffe (1971), Crowston et al. (1973), and Hartung (1973). The specification of demand parameters and price variations due to any retail promotional strategies is exogenous to the optimization model, hence does not affect the linearity structure.

3.5 Calculating Unit Sales

Unit sales volume in period t is bounded by the period's demand, so

$$U_j(t|\xi_1, \xi_f) \leq d_j(t|\xi_f), \quad \text{for all } j, t, \xi_1, \xi_f. \tag{8.10}$$

While traditional inventory models assume that lost sales occur only when inventory is fully exhausted, in retail marketing environments the amount of on-hand inventory can influence sales. In apparel, for example, sales rates can deteriorate as inventory drops because the remaining inventory consists of increasingly broken assortments with incomplete selections of sizes and colors (Smith and Achabal (1998)). Low inventory also increases the likelihood that some stores are inadequately stocked, i.e., the inventory is not "balanced." While the relationship between inventory level and sales is not necessarily linear (Smith and Achabal (1998)), a linear approximation is reasonable within the range of values of the inventory level that is expected in practice. This lends considerable analytical tractability to our formulation. Therefore, we allow unit sales to depend upon the beginning inventory according to the following constraints:

$$U_j\left(t|\xi_1,\xi_f\right) \le \eta_j I_j\left(t|\xi_1,\xi_f\right), \text{ for all } j,t,\xi_1,\xi_f \tag{8.11}$$

where $\eta_j \equiv$ maximum fraction of the beginning inventory that can be sold in one period[11]. Because of (8.1) and the production capacity constraints in (8.6) and (8.7), it is also possible that neither (8.10) or (8.11) will be binding for a given *t*.

Constraints (8.10) and (8.11) assume that the unfilled demand is lost (to competitors, for example), which is more common than backordering for most retail merchandise. Backordering, which is actually more straightforward to model, can easily be accommodated within our formulation by modifying the inventory balance equations.

3.6 The Objective Function

The objective function will be defined in terms of the following economic parameters:

$\pi_j(t) \equiv$ average selling price for product j in period t

$c_{ij} \equiv$ unit procurement + shipping cost ("landed cost") for product j purchased from vendor i

$r_j \equiv$ residual value per unit of product j at the end of the selling season

$c_j^F \equiv$ cost per yard of fabric for product j

$r_j^F \equiv$ residual value per yard of fabric for product j at the end of the selling season

$h_j \equiv$ retailer's unit holding cost per period for product j

[11] Retailers typically track the "sell-through" rate, i.e., the fraction of the beginning on-hand inventory that is sold in each time period. If the sell-through rate is too high, it is assumed that some sales have been lost due to insufficient inventory (see Smith et al. (1998) for further discussion).

$v_{ij} \equiv$ vendor i's unit storage charge per period for product j

The average selling price $\pi_j(t)$ may vary by time period to allow periodic price markdowns during the season. The value of r_j has different interpretations for seasonal and fashion items. For a seasonal item, it corresponds to the unit value of this product in the next selling season (i.e., the avoided replacement cost minus any holding cost). For fashion items it describes a "salvage value." At the selling season's end, any remaining fashion items may be sold through outlet stores or in bulk to discounters, resulting in markdowns to prices possibly below the original cost.

The expected revenue and cost for each product, denoted as R_j and C_j, respectively, are:

$$\begin{aligned} R_j &= \sum_{t,\xi_1,\xi_f} p(\xi_1,\xi_f)\Big\{\pi_j(t|\xi_f)\,U_j(t|\xi_1,\xi_f) \\ &\quad + r_j I_j(t_f|\xi_1,\xi_f) + r_j^F\left(F_j - Z_j^F(\xi_1)\right)\Big\} \qquad (8.12) \\ C_j &= \sum_{i,t,\xi_1} p(\xi_1)\{c_{ij}P_{ij}(t|\xi_{ij}) + v_{ij}M_{ij}(t|\xi_1)\} \\ &\quad + \sum_{t,\xi_1,\xi_f} p(\xi_1,\xi_f)\,h_j I_j(t|\xi_1,\xi_f) + c_j^F F_j \qquad (8.13) \end{aligned}$$

where $p(\xi_1,\xi_f)$ and $p(\xi_1)$ are the previously defined joint and marginal probabilities, respectively. The total objective to maximize is then $\sum_j \{R_j - C_j\}$

The fabric commitments, production capacity commitments, and shipping schedules that optimize this objective function correspond to a sequence of decisions under uncertainty, where the demand information changes at each decision point. In general, this can be viewed as a stochastic dynamic programming problem (with linear constraints). Unfortunately, the size of the resulting state space and the complexity of the objective make this solution approach impractical. However, as long as the states of information are restricted to a discrete set of values, the equations for R_j and C_j are linear in the decision variables, so that this optimization problem is a linear program[12].

[12]This approach for handling uncertainty within an LP formulation was first suggested by Dantzig (1955). Including decision variables whose values may be chosen after the resolution of the uncertainty leads to what is generally termed as a stochastic linear program with recourse. See Hansotia (1980) and Infanger (1994) for discussion of various technical aspects of solving such models and extensive reviews of the literature.

3.7 Model Extensions for Sensitivity Analysis

Important insights from an optimization analysis are often derived from shadow prices and other sensitivity outputs. In vendor sourcing, this information can identify the most critical vendor production and storage constraints, and therefore guide the retailer in negotiating these limits or in identifying alternative vendors with appropriate capabilities. The retailer's storage limits or end–season inventory requirements may also be opportunities for performance improvement.

Because of the multitude of variables and constraints associated with the specific time periods and information states, most individual shadow prices in our model are not directly meaningful. However, useful sensitivity information can be obtained by introducing additional variables. For instance, since increases in production and storage capacity would typically be made for the entire horizon rather than by individual periods, it is appropriate to introduce a single variable that increments a given vendor's capacity uniformly in all periods and information states. If this variable is then constrained to be 0, the corresponding shadow price will reveal the marginal benefit of increasing the vendor's capacities in all periods at once. We add variables for these aggregate constraints as follows:

$\overline{\Delta}_i \equiv$ increase in quarterly production capacity (000's) for vendor i for all quarters

$\underline{\Delta}_i \equiv$ decrease in quarterly minimum production (000's) for vendor i for all quarters

$\omega_i \equiv$ increase in storage capacity at vendor i (cartons)

The appropriate constraint equations ((8.6) and (8.3)) are then replaced with the following:

$$\underline{k}_i(q) - \underline{\Delta}_i \leq Z_i(q|\xi_1) \leq \overline{k}_i(q) + \overline{\Delta}_i, \text{ for all } i, q, \xi_1 \tag{8.14}$$

$$\sum_j \nu_j M_{ij}(t|\xi_1) \leq w_i(t) + \omega_i, \text{ for all } i, t, \xi_1 \tag{8.15}$$

$$\underline{\Delta}_i, \overline{\Delta}_i, \omega_i = 0. \tag{8.16}$$

This enhancement was made for components of the formulation deemed most important by the retail planners: vendor production capacity, vendor flexibility, vendor storage, end-season retail inventory, and product demand.

3.8 Positioning This Model in the Retailer's Planning Process

Our discussions with executives at our retail sponsor highlighted two key issues relevant to the implementation of our methodology. The first deals with the timing of the analysis. Even though our planning model formulates the demand and supply dynamics over a finite horizon, like many other such models it would actually be used on a rolling horizon basis. (As noted earlier, this approach can be termed "open-loop feedback control.") Thus, the production planning actions recommended by each run of the model will serve as important inputs to the subsequent run[13]. The second issue deals with the level of product aggregation at which the analysis is performed. The retail executives envisioned this model being used for analysis at the product category level (e.g., T-shirts, denim jackets, or denim pants) as well as at a lower product type level (e.g., Pocket Tees, V-Neck Tees, and Crew Neck Tees). The former analysis will typically be of interest to product managers who are responsible for the profitability of separate categories. The latter will be of primary interest to buyers who devise procurement plans for product types.

4. The Decision Support System

With extensive input from sourcing managers at the retail chain, the optimization model described above was implemented as a PC-based decision support system (DSS) named the *Sourcing Allocation Manager* (SAM). The user interface screens were programmed in Visual Basic and the optimization engine is LINGO, supplied to us by LINDO Systems. For test problems with four products, four vendors, a nine-month planning horizon, and 27 distinct sample paths of information realizations, the LP has several thousand decision variables and constraints. It was solved on a 300 MHz Pentium II PC in approximately 3-5 minutes.

The DSS development was a "proof of concept" exercise with several goals: (1) to provide a context for defining the user inputs and outputs of the model, (2) to test the practical viability of the optimization algorithm, (3) to demonstrate to the sourcing managers the potential benefits of the system, and (4) to identify through experience the cost tradeoffs

[13]For example, within the context of our formulation, at time t_0 one could be planning for a six month selling season that begins six months hence (i.e., $t_b - t_0 = 6$ months, and $t_f - t_b = 6$ months). The entire planning horizon thus consists of 4 quarters, with planning decisions being revised at a weekly level. In this case, the previously committed production, which might be the result of a prior run of the model, can serve as input constraints to the current run of the model.

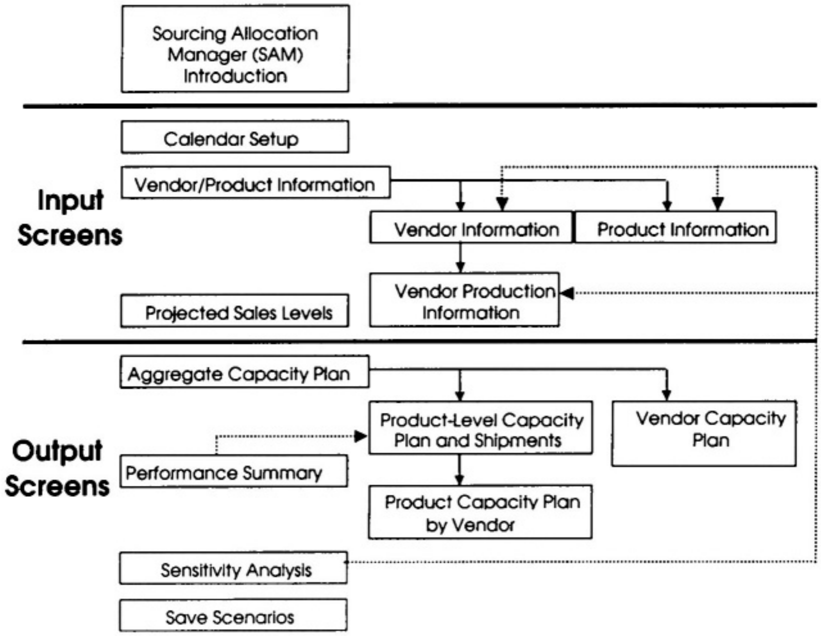


Figure 8.4. SAM Screen Flowchart

most important for planning. These goals were largely achieved, and the extensive involvement of retail planners profoundly influenced the resulting system in numerous ways. For instance, over the course of its development, the DSS evolved from a batch-processing application that generated the optimal sourcing plan for a particular text file consisting of all relevant parameters, to a system that enables beginning users to perform sensitivity and cost tradeoff analyses interactively. Feedback from users over the course of their cumulative experience with the DSS led to a number of key enhancements of the core mathematical formulation as well.

4.1 Graphical User Interface

The logical flow of the DSS screens is illustrated in Figure 8.4. In general, all input screens must be completed before any output screens can be viewed, although input scenarios can be stored for subsequent analyses. Data for the input screens can either be keyed in manually or read from a Microsoft Excel spreadsheet file, which a user can view and modify interactively. In a full-scale implementation most of these values would likely be fed directly from other applications or databases.

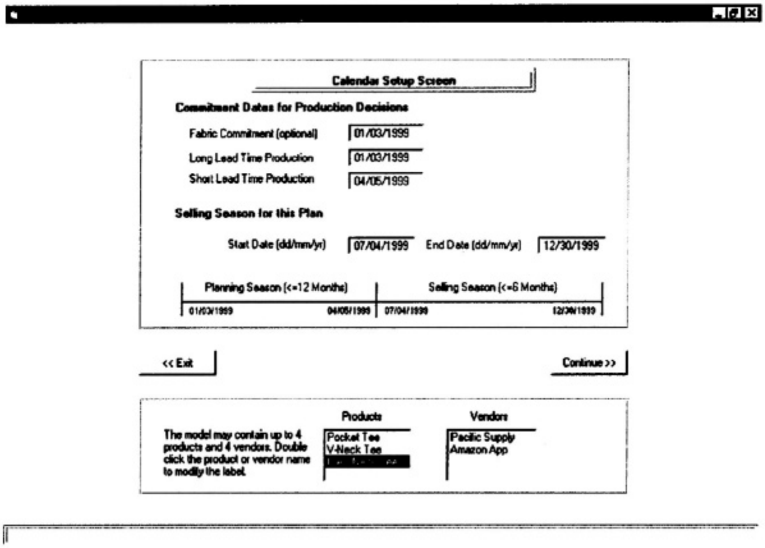


Figure 8.5. Calendar Setup Screen

Below we will describe the main screens, although space limitations preclude the inclusion of all screen views.

Input Screens

The Calendar Setup Screen in Figure 8.5 allows the user to specify dates delineating the timetable for planning. The first date field is for Fabric Commitment, indicating when the fabric must be ordered for all products under consideration. The second and third dates are the commitment times for the long and short lead-time vendors, respectively. For reasons discussed in Section 3.1 and Section 3.2, all commitments with vendors are modeled as being made at one of these two dates. However, a single vendor is allowed different commitment dates for different products. The Selling Season corresponds to the retailer's season for this set of products or the time frame for which this set of production commitments is in effect, whichever is shorter. (For continuing products, linkage to selling periods beyond this season is achieved by requiring end of season inventory, as described in Section 3.3.) This screen also allows the specification of vendor and product names (up to 4 of each).

Figure 8.6 shows the screen displaying vendor and product attributes. These are organized into a matrix with column headings (vendor names), row headings (product names), and interior cells that each provide click-

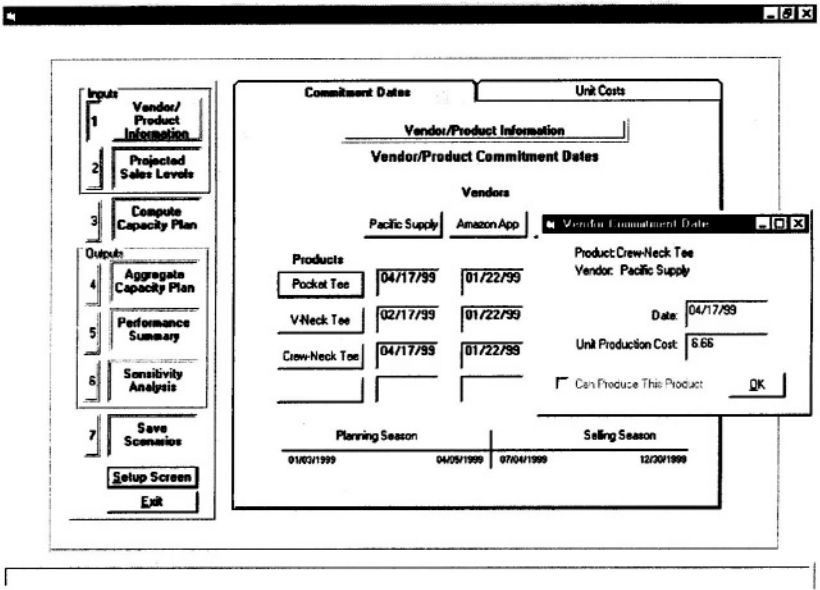


Figure 8.6. Vendor/Product Information Screen

through access to the appropriate type of information. This matrix framework persists throughout the DSS.

This screen has two different views accessible via the folder "tabs" at the top of the screen. These present the most significant attributes of each vendor-product sourcing combination – Commitment Date and Unit Cost. In the Commitment Date view depicted in Figure 8.6, the interior cells report the commitment deadlines required by each vendor for supplying each product. Short and long lead-time vendors are differentiated by color coding of these cells (although this is not apparent in a black-and-white graphic). Clicking on an interior cell calls up the dialog box presented in the foreground, in which a user can view or alter the Commitment Date or the Unit Cost. Toggling to the Unit Cost view presents the matrix of unit costs for all vendor-product combinations. Reflecting the richness of detail which our model can accommodate, separate input screens are required to fully specify the attributes of each vendor and each product. These may be accessed by clicking the appropriate column or row heading buttons, as described below.

Clicking a column heading button in the Figure 8.6 screen calls up the vendor information shown in Figure 8.7. The allowable quarter-to-quarter production volume adjustment (see equation (8.7)) and shipping

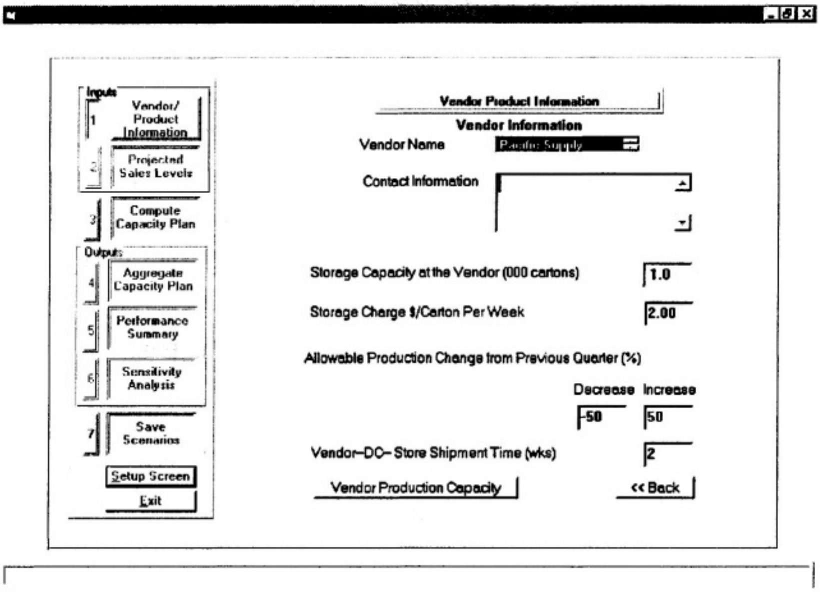


Figure 8.7. Vendor Information Screen

lead time dictate the relative flexibility of this vendor. The Vendor Production Capacity button allows access to a screen detailing each vendor's total production capacity by quarter, shown in Figure 8.8. The storage capacity and total quarterly production capacity are shared across all products made by this vendor.

Clicking a row heading button in the Figure 8.6 screen calls up the Product Information screen shown in Figure 8.9. Here each product's sales forecasts (for the "Most Likely" case, as described in Section 3.1), retail prices (week by week to accommodate frequent price changes if dictated by the retailer's promotional strategy), inventory costs and requirements, and fabric information are entered (or taken from a spreadsheet using an embedded interface accessible from the "Show Spreadsheet" button) and displayed. The inventory constraints and costs on this screen apply only to inventory held in the retailer's distribution system and stores.

The Projected Sales Levels button on the left-hand menu calls up the screen shown in Figure 8.10, which solicits the retail planners' beliefs about demand uncertainty. The "Most Likely" total season forecast for each product is automatically computed by summing the estimated weekly sales shown in Figure 8.9. The user specifies what a "Low"

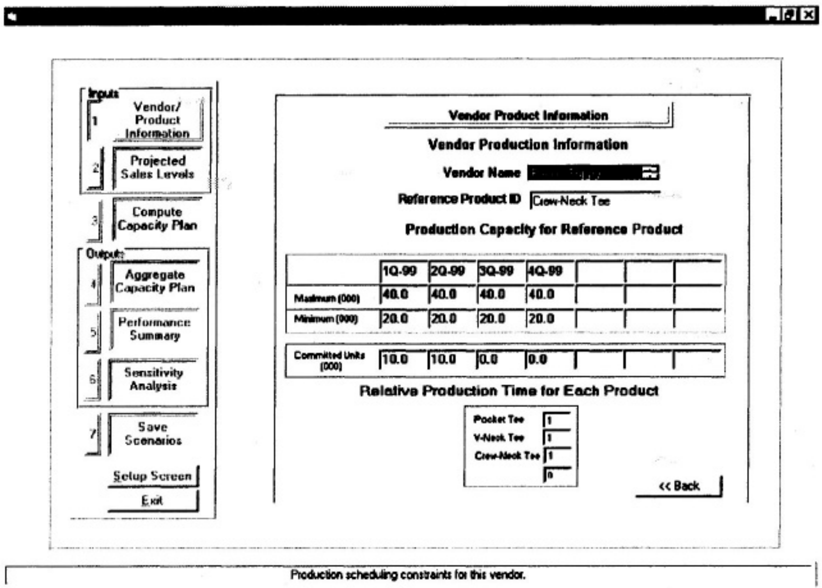


Figure 8.8. Vendor Production Information Screen

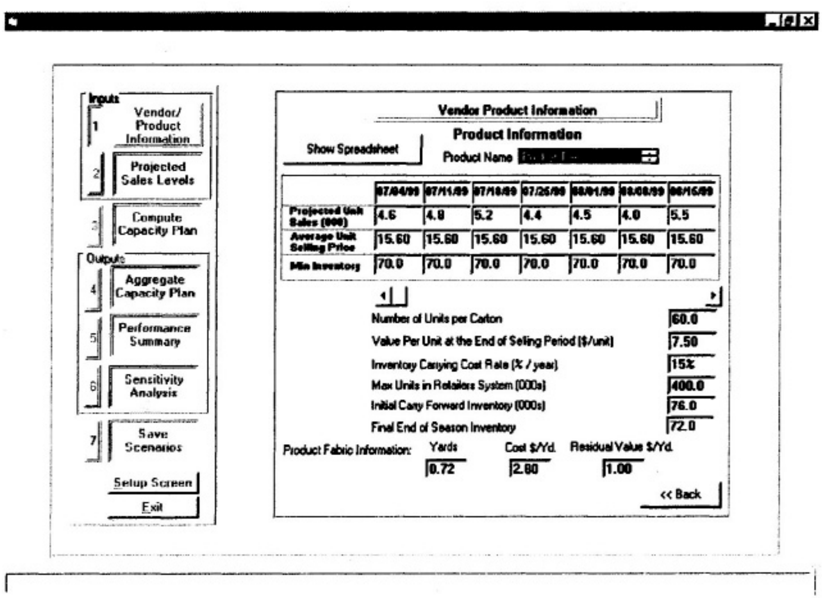


Figure 8.9. Product Information Screen

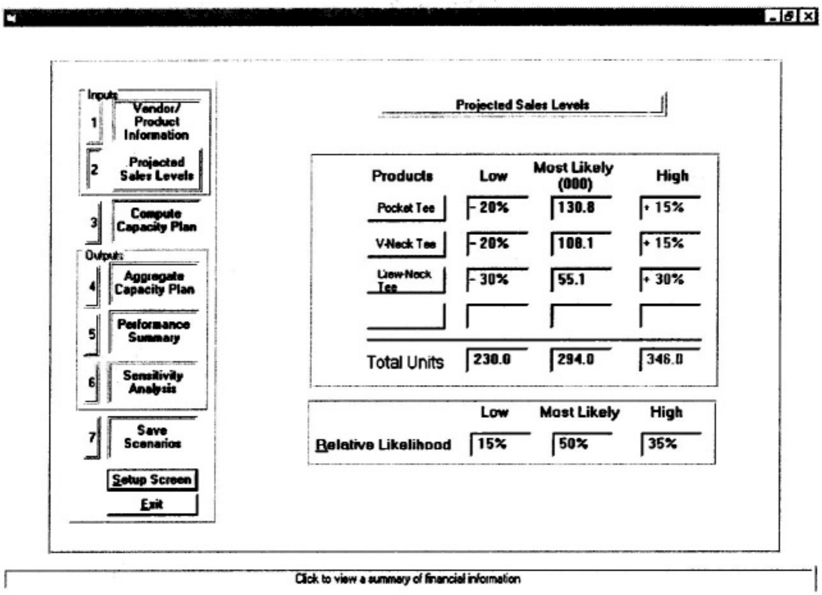


Figure 8.10. Projected Sales Levels Screen

and "High" forecast update would mean for each product in terms of a percentage deviation from the "Most Likely" volume. The percentage changes input here for a product capture the uncertainty about its demand, i.e., the extent to which the projections about that product's demand might change between t_0 and t_1. Stable products will tend to have more narrow ranges than newer or fashion-oriented products. These parameters are used to scale the weekly sales according to the demand model described in Section 3.4. At the bottom area of the screen the user must specify relative likelihoods for each of the three scenarios. After considerable discussion and experimentation, this input format was preferred by the sourcing managers because they are accustomed to developing strategies for three scenarios (cf. footnote 7).

Output Screens

A complete set of inputs allows the optimal sourcing plan to be determined by the LP solver engine. Since this plan contains considerable detail as well as contingency plans, the output is summarized across several screens. The main output screen is shown in Figure 8.11, which reports the total amount of each product that should be committed to each vendor under the three scenarios. (The buttons along the bottom

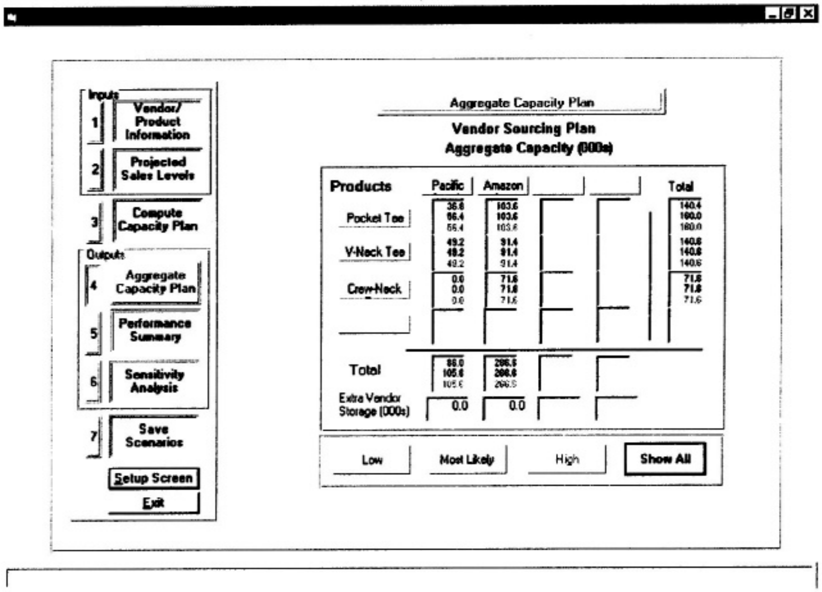


Figure 8.11. Aggregate Capacity Plan

of the screen allow the user to toggle through the plans for each individual scenario, or to juxtapose all three as shown in the figure. Each scenario button has a different color, which is used to display the corresponding plan in the "Show All" view.) By definition, long lead-time vendors must have the same commitments in all scenarios, while short lead-time vendors may receive commitments that depend on the scenario. This flexibility justifies any cost premium the latter vendors may charge. The summary provided by this screen can give each vendor a reasonable picture of how its total volume of business might vary. The buttons on the row and column headings are analogous to those in Figure 8.6, in that they provide paths to further details by vendor or product.

To eliminate any potential LP infeasibility due to the vendor storage constraints, the formulation was modified to allow unlimited auxiliary storage (at some very high price, to discourage the pursuit of this option). The Extra Vendor Storage cells at the bottom of the screen report the additional space (in thousands of cartons) required by the modified formulation's optimal plan. The sourcing managers considered this relaxation to be reasonable since, in spite of formally stated vendor storage limits, additional storage can almost always be obtained at some price.

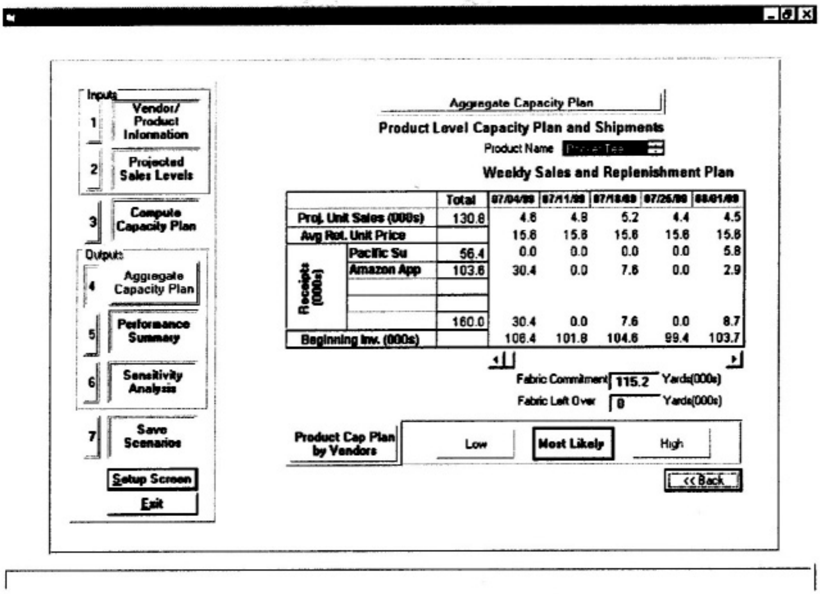


Figure 8.12. Product-Level Capacity Plan and Shipments Screen

The Product-Level Capacity Plan and Shipments Screen in Figure 8.12 shows week-by-week sourcing plans by product, including shipping receipts and fabric commitments for each scenario (displayed, as always, by toggling via the buttons near the screen's bottom). Figure 8.13 provides the vendor perspective on quarterly production and shipments for all products under each scenario. Barcharts are presented beneath the numerical table to visually illustrate the changes in vendor commitments for each product across the scenarios.

Figure 8.14 presents a summary of key retail performance metrics for each of the sales scenarios. GM represents gross margin dollars, inventory turnover is the total annualized sales divided by the average inventory level during the season, and GMROI (Gross Margin Return On Investment) is the product of the inventory turnover and the gross margin per unit (cf. Berman and Evans (1998)). Average cost is scenario-dependent because salvage revenues vary with the amount of residual fabric and end-of-season inventory. Again, the metrics take different values for each scenario, and the barcharts compare the values for each metric across the three scenarios.

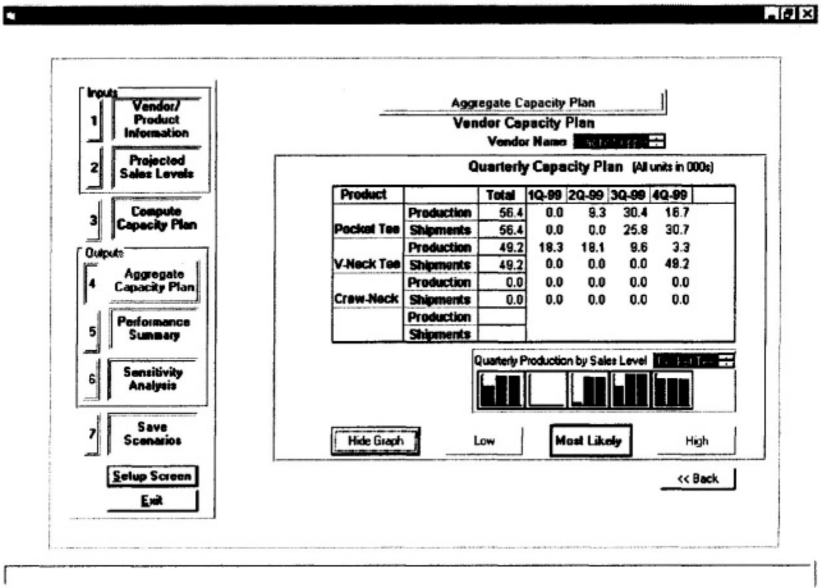


Figure 8.13. Vendor Capacity Plan Screen

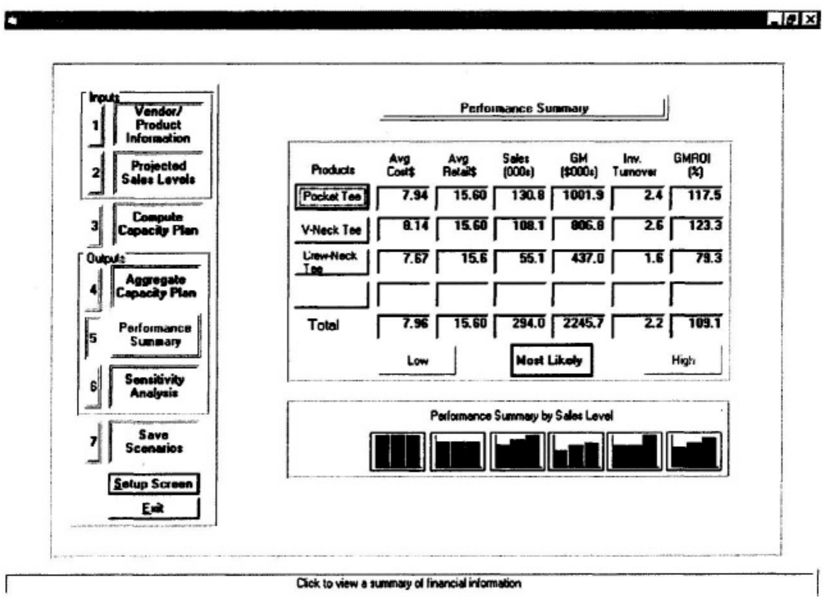


Figure 8.14. Performance Summary Screen

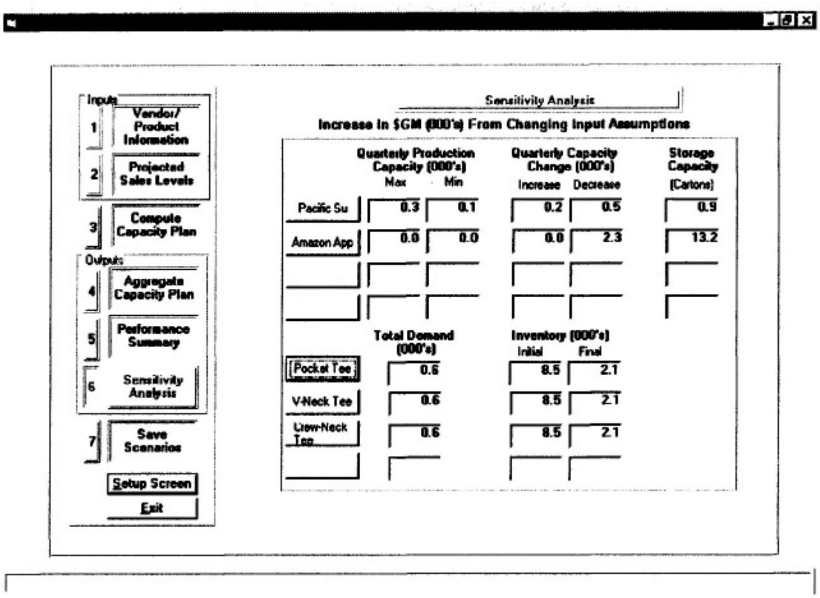


Figure 8.15. Sensitivity Analysis Screen

Sensitivity Analysis

Figure 8.15 shows the screen interface to the sensitivity analysis framework described in Section 3.7. Clicking on the vendor or product name buttons takes the user back to the appropriate input screen to view the current parameter value and make any desired modifications. This allows "What if?" analyses to be conducted quickly and easily for a wide range of assumptions. Added at the request of our retail sponsor, this screen has had the longest evolutionary path of any of the SAM screens and is considered to have the greatest strategic importance. By running the DSS, the retailer can determine which of each vendor's constraints are the most significant obstacles to improved profitability and can also estimate the value of relaxing any particular constraint. This can play a valuable role in negotiating with vendors by providing information that was not previously quantifiable.

Figure 8.16 shows the screen associated with the Save Scenarios button on the left menu. Because of the large number of input parameters, and to facilitate what-if analysis, SAM allows the capture of inputs and outputs for up to five problem scenarios. The user can annotate these in the text window near the bottom of the screen, where the various scenarios can be saved and retrieved.

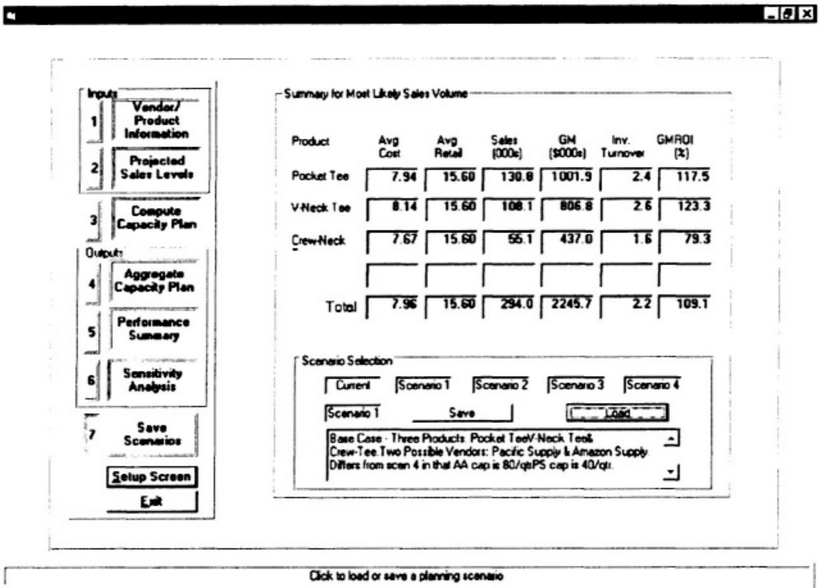


Figure 8.16. Save Scenarios Screen

4.2 User Experiences with the DSS

Experience with the SAM DSS was obtained through an analysis conducted with our sponsor firm, using representative but disguised data. The retailer's goal for this analysis was to gain experience with the model and develop an understanding of the key tradeoffs between vendor capabilities and unit costs. The details are presented as a case study in Agrawal et al. (2001). Some of the resulting insights are as follows:

1 Building a stochastic model allows presentation of distributional information about any system performance metric. This can provide valuable insight about the extent of intrinsic risk to which a decision-maker is exposed. The sourcing managers were able to connect this to their strategies for managing risk, including those related to the selection of product assortments. This would not be possible under any deterministic planning methodology.

2 Under reasonable assumptions, SAM's recommendations can improve expected profits by several percentage points relative to typical sourcing practices. Since net profit margins in retailing tend to be very small, this suggests that our methodology can offer very meaningful gains.

3 There can be value in using a portfolio of vendors with differing production flexibility. In practice, while buyers know at an intuitive level that flexibility has value, the inability to quantify this has left them biased toward vendors quoting the lowest unit costs. Our model easily demonstrates that additional vendor flexibility can indeed be worth a price premium when demand uncertainty exists, and we provide a means to evaluate this tradeoff with a realistic level of detail.

4 Not all production capacity is created equal. Capacity cannot be properly valued independently of its flexibility constraints, such as commitment lead time and allowable production change from quarter to quarter. The effect of these conditions is a function of the attributes of the type of merchandise, in particular the predictability of demand and the cost of obsolescence.

5 While conventional wisdom suggests that inventory turnover is determined by the replenishment policies adopted at the store level, tension between demand seasonality and the vendors' desire to maintain stable production schedules profoundly affects retailer inventory levels. Thus, efforts to increase turnover should also consider negotiations with vendors to seek greater production flexibility.

6 From an organizational point of view, our methodology can provide a vehicle for facilitating cross-functional communication and negotiation. Specifically, in a retail firm the merchandising, sourcing, and finance organizations typically have somewhat conflicting objectives with respect to inventory management strategy. (In mathematical terms, each group typically perceives a different segment of the overall objective function.) An early insight for us and our corporate sponsors was that our DSS could serve as a tool for brokering the concerns of these groups by solving the global optimization problem, explicitly quantifying tradeoffs, and, most importantly, defining a common vocabulary for discussion.

5. Conclusion

Estimating the value of adding or dropping a vendor, renegotiating the terms of a supply contract, or improving forecast capability requires the respecification of the production schedule in ways that may differ dramatically from past plans. The complexity of such decisions renders the subjective selection of optimal or even near-optimal plans extremely difficult or impossible. While many retail buyers and merchandise planners rely on extensive databases and query tools for decision support, there are few computer-based methods for optimal decision making or sensitivity analysis regarding these decisions. Our model and the associated decision support software provide retail planners with the power to identify and evaluate a wide variety of potential supply chain improvements that they are not currently able to consider.

Capturing market uncertainty through discrete scenarios is a familiar mechanism that simplifies the required user inputs and allows the application of linear programming optimization. Because of the many types of production and sales constraints that may apply in a retail environment, simplicity of use is essential to the practicality of a decision support tool. Tests of our model by buyers and planners within a major retail organization indicate that our framework is compatible with the production commitment decisions they face.

Acknowledgments

The authors are grateful to the corporate sponsors of the Retail Workbench at Santa Clara University for their financial support, for suggesting the business problem, and for their comments and suggestions in the course of this research. We are especially grateful to Professor Dale Achabal, Retail Workbench Director, who provided valuable guidance throughout the project, and to executives from the retail firm sponsoring this project for working with us in developing the modeling assumptions and designing the user screens for the software tool. The software owes much of its existence to Jerry Currie, who wrote and debugged the hundreds of lines of Visual Basic code that went into it. We would also like to acknowledge LINDO Systems for providing a special version of its LINGO optimization software for integration into our system. The authors are fully responsible for the opinions expressed and any remaining errors.

References

Achabal, D.D., S.H. McIntyre, and S.A. Smith (1990). Maximizing Profits from Periodic Department Store Promotions. *Journal of Retailing,* 66, 4, 383-407.

Agrawal, N., S.A. Smith, and A.A. Tsay (2001). Multi-Vendor Sourcing in a Retail Supply Chain. Working Paper, Leavey School of Business, Santa Clara University, Forthcoming, *Production & Operations Management.*

Berman, B. and J.R. Evans (1998). *Retail Management: A Strategic Approach,* Prentice Hall, Upper Saddle River, NJ.

Bertsekas, D.P. (1976). *Dynamic Programming and Stochastic Control,* Academic Press, New York, NY.

Bitran, G.R., E.A. Haas, and H. Matsuo (1986). Production Planning of Style Goods with High Setup Costs and Forecast Revisions. *Operations Research,* 34, 2, 226-236.

Brown, A.O. and H.L. Lee (1998). Optimal 'Pay to Delay' Capacity Reservation with Application to the Semi-conductor Industry. Working Paper, Owen Graduate School of Management, Vanderbilt University, Nashville, TN.

Buxey, G. (1993). Production Planning and Scheduling for Seasonal Demand. *International Journal of Operations & Production Management,* 13, 7, 4-21.

Buxey, G. (1995). A Managerial Perspective on Aggregate Planning. *International Journal of Production Economics,* 41, 1-3, 127-133.

Chang, S.H. and D.E. Fyffe (1971). Estimation of Forecast Errors for Seasonal Style-Goods Sales. *Management Science,* 18, 2, B89-B96.

Cox, J.C. and M. Rubinstein (1985). *Options Markets,* Prentice-Hall, Englewood Cliffs, NJ.

Crowston, W.B., W.H. Hausman, and W.R. Kampe (1973). Multistage Production For Stochastic Seasonal Demand. *Management Science,* 19, 8, 924-935.

Dantzig, G.B. (1955). Linear Programming Under Uncertainty. *Management Science,* 1, 3, 197-206.

Donohue, K.L. (2000). Efficient Supply Contracts for Fashion Goods with Forecast Updating and Two Production Modes. *Management Science,* 46, 11, 1397-1411.

Eppen, G.D. and A.V. Iyer (1997). Backup Agreements in Fashion Buying - The Value of Upstream Flexibility. *Management Science,* 43, 11, 1469-1484.

Eppen, G.D., R.K. Martin, and L. Schrage (1989). A Scenario Approach to Capacity Planning. *Operations Research,* 37, 4, 517-527.

Fisher, M. and A. Raman (1996). Reducing the Cost of Demand Uncertainty Through Accurate Response to Early Sales. *Operations Research,* 44, 1, 87-99.

Guerrero, H.H., K.R. Baker, and M.H. Southard (1986). The Dynamics of Hedging the Master Schedule. *International Journal of Production Research,* 24, 6, 1475-1483.

Gunther, H.O. (1982). A Comparison of Two Classes of Aggregate Production Planning Models Under Stochastic Demand. *Engineering Costs and Production Economics,* 6, 89-97.

Hansotia, B.J. (1980). Stochastic Linear Programming With Recourse: A Tutorial. *Decision Sciences,* 11, 1, 151-168.

Hartung, P.H. (1973). A Simple Style Goods Inventory Model. *Management Science,* 19, 12, 1452-1458.

Hausman, W.H. and R. Peterson (1972). Multiproduct Production Scheduling For Style Goods With Limited Capacity, Forecast Revisions And Terminal Delivery. *Management Science,* 18, 7, 370-383.

Hausman, W.H. and R. Sides (1973). Mail-order Demands For Style Goods - Theory And Data Analysis. *Management Science,* 20, 2, 191-202.

Heath, D.C. and P.L. Jackson (1994). Modeling The Evolution Of Demand Forecasts With Application To Safety Stock Analysis In Production/Distribution Systems. *IIE Transactions,* 26, 3, 17-30.

Hertz, D.B. and K.H. Schaffir (1960). A Forecasting Method for Management of Seasonal Style-Goods Inventories. *Operations Research,* 8, 1, 45-52.

Huchzermeier, A. and M. A. Cohen (1996). Valuing Operational Flexibility Under Exchange Rate Risk. *Operations Research,* 44, 1, 100-113.

Hunter, N.A., R.E. King, and H.L.W. Nuttle (1992). An Apparel-supply System for QR Retailing. *Journal of the Textiles Institute,* 83, 3, 462-471.

Hunter, N.A., R.E. King, and H.L.W. Nuttle (1996). Evaluation of Traditional and Quick-response Retailing Procedures by Using a Stochastic Simulation Model. *Journal of the Textiles Institute,* 87, 1, 42-55.

Infanger, G. (1994). *Planning Under Uncertainty: Solving Large-Scale Stochastic Linear Programs,* Boyd & Fraser, Danvers, MA.

Kalyanam, K. (1996). Pricing Decisions Under Demand Uncertainty: A Bayesian Mixture Model Approach. *Marketing Science,* 15, 3, 207-221.

King, R.E. and N.A. Hunter (1996). Demand Re-estimation and Inventory Replenishment of Basic Apparel in a Specialty Retail Chain. *Journal of the Textiles Institute,* 87, 1, 31-41.

Kira, D., M. Kusy, and I. Rakita (1997). A Stochastic Linear Programming Approach to Hierarchical Production Planning. *Journal of the Operational Research Society,* 48, 2, 207-211.

Miller, J.G. (1979). Hedging the Master Schedule. In *Disaggregation Problems In Manufacturing And Service Organizations.* L.P. Ritzman, et al. (Eds.) Martinus Nifhoff, Boston, MA, 237-256.

Murray, G.R. and E.A. Silver (1966). A Bayesian Analysis of the Style Goods Inventory Problem. *Management Science,* 12, 11, 785-797.

Nam, S. and R. Logendran (1992). Aggregate Production Planning–A Survey of Models and Methodologies. *European Journal of Operational Research,* 61, 3, 255-272.

Nuttle, H.L.W, R.E. King, and N.A. Hunter (1991). A Stochastic Model of the Apparel Retailing Process for Seasonal Apparel. *Journal of the Textiles Institute,* 82, 2, 247-259.

Ravindran, A. (1972). Management of Seasonal Style-Goods Inventory. *Operations Research,* 20, 2, 265-275.

Riggs, W.E. (1984). A Short-Term Forecasting Model for Producers of Seasonal Style Goods. *Production & Inventory Management,* 25, 4, 42-49.

Riter, C. (1967). The Merchandising Decision Under Uncertainty. *Journal of Marketing,* 31, 44-47.

Robison, L.J. and P.J. Barry (1987). *The Competitive Firm's Response to Risk,* Macmillan, New York, NY.

Silver, E.A. and R. Peterson (1985). *Decision Systems for Inventory Management and Production Planning,* 2nd Edition, Wiley, New York.

Smith, S.A. and D.D. Achabal (1998). Clearance Pricing and Inventory Policies for Retail Chains. *Management Science,* 44, 3, 285-300.

Smith, S.A., N. Agrawal, and S.H. McIntyre (1998). A Discrete Optimization Model for Seasonal Merchandise Planning. *Journal of Retailing,* 74, 2, 193-221.

Wadsworth, G.P. (1959). In *Notes on Operations Research,* assembled by the Operations Research Center, M.I.T., Technology Press, Cambridge, MA.

Wolfe, H.B. (1968). A Model for Control of Style Merchandise. *Industrial Management Review,* 9, 2, 69-82.

Chapter 9

COORDINATING THE DISTRIBUTION CHAIN: NEW MODELS FOR NEW CHALLENGES

Anantaram Balakrishnan
The Smeal College of Business Administration
The Pennsylvania State University
University Park, PA 16802
anantb@psu.edu

Joseph Geunes
Department of Industrial and Systems Engineering
The University of Florida
Gainesville, FL 32611
geunes@ise.ufl.edu

Michael S. Pangburn
The Smeal College of Business Administration
The Pennsylvania State University
University Park, PA 16802
mikepangburn@psu.edu

Abstract Recent business and technological trends have transformed the structure and performance requirements for distribution channels in many industries. Higher service level expectations of retail customers, distribution outsourcing by manufacturers, and the proliferation of advanced information technologies drive these transformations, presenting new problems in supply chain management. Motivated by a project with a leading building-products manufacturer, this paper addresses some of these new issues. Over the past two decades, this manufacturer witnessed the migration of building-products sales from independent specialty retailers to large retail chains, prompting it to create a new

J. Geunes et al. (eds.),
Supply Chain Management: Models, Applications, and Research Directions, 199-242.

network of independent distributors to meet the service expectations of these 'big-box' retailers. This paper addresses three important challenges in managing the new distribution network. We first develop a fee-setting model to decide the manufacturer's compensation scheme for the services provided by its independent distributors. Next, we address a tactical distribution planning problem, incorporating resource acquisition and deployment decisions, for scheduled deliveries when demand is highly variable. Lastly, we investigate possible mechanisms for limiting retail store-order variability, and analyze the system-wide cost benefits resulting from variability reduction. In addition to identifying new modeling opportunities and discussing their implications for the building-products manufacturer, this paper highlights new research opportunities resulting from the evolving dynamics of supply chain management.

1. Introduction

Distribution channels in many industries have experienced major transformations in recent years – in their architecture, collaborative partnerships, operations practices, and performance requirements. These channel transformations stem from several factors that have altered the ground rules for providing competitive distribution services. We attribute these changes largely to three major driving forces:

1 *Customers and retailers are raising their suppliers' service expectations.* Customer demands for product variety and ready availability are propagating throughout supply networks. Powerful retail chains, responding to these customer needs, have in turn increased their service expectations, requiring that their suppliers provide shorter delivery lead times and more frequent deliveries of a wider variety of products in smaller batches.

2 *Manufacturers are increasingly outsourcing distribution.* Partnerships with third party logistics providers allow manufacturers to focus on their core competencies while taking advantage of the distribution efficiency and expertise of independent distributors. In turn, distributors are expanding the scope of their activities beyond their traditional warehousing and transportation functions to include customization and postponement services (such as merge-in-transit and just-in-time assembly).

3 *Information technologies are providing more timely and detailed supply chain data.* Advances in information technologies–both in reach and connectivity–increase the potential for information sharing and enable tighter integration among channel partners.

New web-based and enterprise integration systems permit visibility of information from various points in the supply chain, significantly reduce information and order lead times, and facilitate distributed monitoring and control.

These transformations present many new planning and decision making challenges. Recent supply chain management research identifies many modeling opportunities associated with these challenges. For instance, studies of optimal buyer-supplier contracting schemes specify appropriate service level requirements, revenue-sharing incentives, and delivery and payment terms (e.g., Tsay, 1999, Li and Kouvelis, 1999, Moinzadeh and Nahmias, 2000). A significant body of work focuses on multi-stage supply chain planning and coordination models. These models typically seek to streamline the flow of goods through the chain by tightly integrating the operations of channel partners or aligning their incentives (e.g., Cachon and Zipkin, 1999, Lee and Whang, 1999, Chen, 1999, Porteus, 2000). Another stream of literature emphasizes the benefits of sharing accurate and timely information in supply chains, made possible through advances in information technology (e.g., Gavirneni, Kapuscinski, and Tayur, 1999, Lee, So, and Tang, 2000). Our aim is to add to the existing literature by discussing three new models–on delivery compensation, distribution planning, and order variability reduction–that support improved distribution chain coordination and performance. These models are motivated by a project with a leading building-products manufacturer who recently redesigned its partnered distribution network to respond to the needs of its important retail customers.

1.1 Project context

Trends over the past decade in the distribution channels for building and home improvement products illustrate the channel transformations many industries have undergone. In the 1970s, the overwhelming majority of building-products sales to the residential market occurred through small, independent hardware stores. The channel transformation for building-products sales began in the 1980s when national home improvement retail chains such as The Home Depot and Lowe's began to grow rapidly in sales and market share. These 'big-box' superstores now dominate the market for do-it-yourself home-improvement products by offering one-stop-shopping for a wide variety of products at low prices. The big-box retailer concept thrives in many industries today (e.g., Sam's Club, Toys R' Us, Office Depot). Major retail chains such as The Home Depot, which has opened over 1000 stores since its founding in 1978, are often successful in leveraging their vast purchasing power to negotiate low prices and strict delivery terms from suppliers. Offering a wide variety of products to consumers with slim profit margins requires big-box stores to keep low inventories; to achieve these goals, big-box stores insist on receiving frequent and reliable deliveries from suppliers under short lead times. Meeting these strict delivery requirements presents significant new challenges to suppliers.

Faced with these challenges, a leading building-products manufacturer (henceforth called the "manufacturer" or "firm") initiated a project with the authors to study and suggest ways to improve its distribution planning practices and operations. The manufacturer supplies over 250 SKUs–grouped into three main product lines–to each of over 1,000 nationwide big-box retail stores and warehouses every week. Stores and warehouses can place their orders up to one day prior to their scheduled delivery date, and are not required to order in large batches. Big-box customers require a single point-of-contact (within the manufacturer's organization) for placing orders, and expect consolidated shipments of items from the manufacturer's various product lines. To meet these needs, the manufacturer established a nationwide network of approximately 10 independent distributors to support distribution to the big-box retailers. Each distributor is responsible for servicing the product needs of all retail stores within its exclusive geographical region. At the time of the study, the manufacturer had already made some overarching strategic logistics decisions, such as the locations of distribution centers and their assigned retail stores. So these issues were outside the scope of our study.

The role of distributors in the big-box network differs significantly from their traditional role. Traditionally, distributors typically carried only a subset of the manufacturer's product lines and were responsible for providing sales and credit services for the independent retailers. In contrast, distributors serving the big-box retailers have no responsibility for sales and credit functions, but they must stock and distribute all of the product lines and supply a large geographic region. For the big-box retailer accounts, the manufacturer handles the marketing and sales of its products, including retail pricing and handling accounts receivables.

Each store in the big-box retail network places its orders directly with the manufacturer, and requires regular deliveries of consolidated shipments containing all ordered items at scheduled times each week. Nearly all stores receive replenishments of the manufacturer's products once per week. Upon receiving an order from a store, the manufacturer transmits the order to the corresponding regional distributor. The distributor is then responsible for delivering the ordered goods at the scheduled delivery time. Each distributor is responsible for managing its own delivery and inventory processes, but must ensure that the ordered products are delivered to each store on time.

To facilitate the coordination and control of the new distribution system, the manufacturer recently made significant investments in upgrading their information technology (IT) capabilities. The new IT systems facilitate information visibility throughout the supply chain. The manufacturer can now track detailed point-of-sale (POS) transactional data from retailers, retailer-to-manufacturer orders, and distributor-to-store shipment data on a daily basis. Our research project with the manufacturer, initiated approximately one year after the new big-box distribution network was established, started with a broad charter: using the detailed transactional data on sales, shipments, and inventories at various locations throughout the big-box supply chain, conduct a "management audit" of current logistics and distribution operations in the big-box network, and identify opportunities to improve the performance of the system. The project was motivated both by the desire to exploit the newly available data, and also by feedback (complaints and suggestions) from distributors and customers.

In this paper, we discuss three specific modeling opportunities our diagnostic analysis identified. The first problem, *fee-setting for delivery services,* occurs at the manufacturer-distributor interface. The manufacturer must establish a consistent, fair, and equitable scheme to compensate its big-box distributors for the warehousing and delivery services they provide. Our initial data analysis indicated that the manufacturer was over-compensating certain distributors (relative to their costs), while under-compensating others. We developed a new delivery fee-setting model that reduces over-compensation while ensuring that no distributor is under-compensated. The second problem, *distribution planning with stochastic demands and scheduled deliveries,* focuses on improving distributor delivery operations and capacity planning. In order to consistently meet delivery commitments, distributors create fixed delivery routes that do not change from week-to-week. In constructing these routes, however, distributors have difficulty planning for the impacts of variability in stores' weekly orders. We developed a distribution planning model that explicitly accounts for demand variability (and its associated costs) when designing delivery routes. Our model considers the availability and costs of "overflow" options for delivering on time when demand exceeds truck capacity. The third problem, *upstream variability control mechanisms,* focuses on improving operations at the distributor-retailer interface. Many distributors felt they were incurring unnecessarily high safety stock and delivery costs due to stores' highly variable order quantities. After verifying that the stores' order variability led to distribution system inefficiencies, we developed an analytical framework encompassing several logistics coordination models to study the ability of certain variability control mechanisms to lower system-wide logistics costs, including inventory, backlog, transportation, and expediting costs.

The three problems we discuss in this paper have not been previously addressed in the literature. We propose appropriate deterministic and stochastic optimization models to address each problem, and apply the models using real data from the manufacturer. Our model formulations are designed to be sufficiently generic to permit application outside of the building-products industry. This paper motivates and develops each model, and summarizes the results from our analysis and computations. The remainder of this paper is organized as follows. Section 2 discusses the fee-setting problem for developing a fair and consistent compensation scheme for distributor deliveries. Section 3 presents our distribution planning model to provide scheduled delivery services when demand is stochastic. Section 4 considers the impacts of order variability on distributor costs, and proposes variability control mechanisms to improve distribution system coordination and performance. Section 5 presents concluding remarks.

2. Optimal Compensation for Distributors

As manufacturers strive to make their supply chains more responsive to market needs, they have come to recognize the critical importance of the distribution function in achieving this goal. Third-party logistics firms or distributors, with their specialized expertise and resources, provide an attractive outsourcing option to handle delivery activities. Just as manufacturers need to develop tight linkages with their vendors, they must also establish collaborative partnerships with distributors to ensure streamlined flow, timely deliveries, and customer satisfaction.

Partnering with third-party distributors requires actions and negotiations on many fronts–installing compatible information systems to exchange transactional data (e.g., conveying orders for delivery), establishing operational procedures and mutual responsibilities, agreeing on service expectations and metrics, and negotiating the payment scheme for compensating distributors for delivery services. In our project with the building-products manufacturer, the process of deciding an appropriate compensation scheme surfaced as an important issue for the firm, one that led to a novel optimization model and decision support system.

We next describe the problem context, formulate a basic version of the problem as a mathematical program, discuss model enhancements, and present results from actual application of the model to the firm's big-box distribution operations. Balakrishnan, Natarajan, and Pangburn (2000) discuss this problem and approach in greater detail.

2.1 Fee-setting decision context

As noted earlier, to supply the customer base consisting of over 1,000 retail stores nationwide, the firm relies on a select group of regional distributors to deliver products from their respective distribution centers (DCs) to assigned stores. The firm compensates distributors for their delivery services by paying a per-delivery fee that depends upon the delivery weight and distance from the DC to the store. This payment scheme, modeled after the tariff structure commonly used in the less-than-truckload (LTL) industry, partitions the store-to-DC distances and delivery weights into (discrete) sets of contiguous distance and weight ranges. For instance, if the maximum DC-to-store distance is 300 miles, the table might categorize distance into three ranges, say, 0 to 50 miles, 50 to 150 miles, and 150 to 300 miles. The fee table contains one row and column, respectively, for each weight and distance range; the value in each cell of this table denotes the fee payable for any delivery that falls into the corresponding distance and weight range. To avoid conflicts and perceptions of differential treatment among distributors, the firm decided to apply the same fee table uniformly to all the distributors. This fee table-based compensation approach offers many advantages. Distributors readily relate to this approach since it is similar to LTL tariff structures, and incorporates the primary drivers of transportation cost: delivery distance and weight. From the manufacturing firm's perspective, the scheme is easy to administer and update (when transportation costs or distribution profiles change), and offers the added advantage of compensating distributors for services rendered rather than simply resources consumed.

Designing the fee table-based compensation scheme entails two sets of decisions: a relatively long-term *range selection* decision, and a periodic *fee setting* decision. Range selection refers to the choice of distance and weight ranges–both the number of ranges, and the width of each range–that define the rows and columns of the fee table. For a given choice of ranges, fee setting (or fee revision) pertains to selecting the fee value for each cell of the corresponding table. Fees are revised periodically (e.g., once a year) due to changes in transportation costs, demand patterns, store assignments, and so on; the ranges impact the structure of the table, and so are changed less often. In this section, we focus on modeling the fee setting decision.

How should the firm select the fee values for each cell of the fee table? One option is to base these fee values on commercial LTL freight rates. This approach has several disadvantages. First, commercial rates vary by carrier and by region, making it difficult to infer a single set of rates to serve as the basis for fee setting. More importantly, commercial rates typically apply to ad hoc shipments, and can therefore be significantly higher than rates that are appropriate for long-term distribution contracts such as those between the firm and its distributors. However, commercial rates do provide some useful benchmarks for fee setting. In particular, these rates and their increments from one range to the next might serve as guidelines (e.g., upper bounds) for selecting fee values.

Although the firm expects to pay its distributors less than commercial rates, it must also ensure that each distributor receives adequate total compensation to cover its cost and provide a reasonable margin. We refer to this condition as the *cost coverage* requirement. Motivated by this consideration, the firm initially (prior to applying our model) used a cost-based approach to decide fee values. The approach consisted of performing a detailed study of delivery routes from DCs to stores assuming deterministic demands, costing each route, and allocating these costs to individual routes in order to estimate the per-delivery costs (and hence fees) for each distance and weight range. Unfortunately, this heuristic method does not necessarily guarantee cost coverage. The inherent difficulties associated with cost allocation schemes combined with the week-to-week variations in actual delivery weights, render these cost estimates unreliable. Indeed, after using the cost-based fee table for a year or so, review of the total actual compensation versus cost for each distributor revealed that some distributors received significant surpluses (i.e., compensation far exceeded costs) whereas others showed marked deficits. Naturally, distributors who experienced deficits argued for increasing the fee values. However, from the manufacturer's perspective, fee increases have a multiplier effect on the firm's total payment, since the same fees apply to *all* distributors, not just those distributors showing a deficit. The firm was, therefore, interested in developing a systematic and "equitable" fee-setting approach that covers distributors' costs while simultaneously reducing disparities in overcompensation.

2.2 Problem definition and formulation

To address the fee-setting decision, we developed a linear programming model that incorporates both market conditions (i.e., commercial rates) and cost considerations via constraints. The model uses data on commercial freight rates to generate upper and lower bounds on permissible fee values. And, instead of allocating costs to individual deliveries, the model employs constraints to ensure that the total expected weekly compensation for each distributor equals or exceeds anticipated costs. We next formally define the fee-setting problem and develop its mathematical formulation.

We are given I weight ranges and J distance ranges that define the rows and columns of the fee table. For $i = 1, 2, \ldots, I$, the i^{th} weight range corresponds to all delivery weights belonging to the pre-specified interval $[w_i, w_{i+1})$; similarly, for $j = 1, 2, \ldots, J$, the given interval $[d_i, d_{i+1})$ defines the distances corresponding to the j^{th} distance range. Suppose the firm uses K distributors, indexed from $k = 1$ to K, to deliver its products. Each distributor k serves a pre-assigned set of stores. For each store, we know the store's distance from its assigned DC (and hence the distance range corresponding to deliveries to this store), and the probability distribution of weekly demand (expressed in terms of total weight of products ordered by the store). Using this information, we can compute, for each distributor, the expected number of deliveries per week that fall within each distance and weight combination in the fee table (see Balakrishnan et al., 2000, for details of this computation). Let a_{ijk} denote this *expected* number of weekly deliveries by distributor k that fall within weight range i and distance range j.

Our model requires an estimate of each distributor's expected total cost per week, denoted as C_k, to deliver orders to its assigned stores. This estimate might include the fixed costs of delivery resources such as trucks and drivers, operational costs for fuel, maintenance and insurance, contingency costs if demand exceeds truck capacities, and a reasonable profit margin. Section 3 describes one approach to estimate this cost using a detailed model of delivery activities; alternatively, the estimate might be based on historical data.

The fee-setting model must decide the fee value for each cell of the table. Let g_{ij} represent the fee value selected for weight range i and distance range j. The fee values must be monotonically increasing, i.e., fees must increase as distance or weight increases. Moreover, to assure acceptance of the fee values by distributors, the firm might wish to relate these values to commercial freight rates by imposing appropriate upper or lower bounds on the individual fee values. Let l_{ij} and u_{ij} denote these exogenous minimum and maximum permitted fee values for cell $(i,\ j)$ of the table. Finally, cost coverage requires that, for each distributor, the total expected compensation must equal or exceed cost. For convenience, we define an auxiliary decision variable s_k denoting the *expected* surplus (i.e., expected compensation minus cost) for each distributor $k = 1, 2, \ldots, K$. The model can be specified using two alternative objective functions. To develop a compensation scheme that is "equitable" across distributors, we might choose to minimize the maximum surplus among all distributors. Alternatively, from the perspective of the manufacturing firm, we might minimize the firm's total expected weekly payments to all distributors, or equivalently minimize the sum of distributor surpluses. Both of these objectives are easy to model. To illustrate the problem formulation, we consider the latter *minsum* objective function, and represent the fee-setting decision problem as the following linear program [FSP]:

[FSP]

$$\textit{minimize} \quad \sum_{k=1}^{K} s_k \tag{9.1}$$

subject to:

$$\textit{Cost Coverage}: \quad \sum_{i=1}^{I}\sum_{j=1}^{J} a_{ijk} g_{ij} - s_k = C_k, \quad k = 1, \ldots, K, \tag{9.2}$$

$$\textit{Monotonicity}: \quad g_{ij} \geq g_{i-1,j}, \quad i = 2, \ldots, I, \; j = 1, \ldots, J, \tag{9.3}$$

$$g_{ij} \geq g_{i,j-1}, \quad i = 1, \ldots, I, \; j = 2, \ldots, J, \tag{9.4}$$

$$\textit{Fee Bounds}: \quad l_{ij} \leq g_{ij} \leq u_{ij}, \quad i = 1, \ldots, I, \; j = 1, \ldots, J, \tag{9.5}$$

$$\textit{Nonnegativity}: \quad g_{ij}, s_k \geq 0, \quad i = 1, \ldots, I, \; j = 1, \ldots, J, \; k = 1, \ldots, K. \tag{9.6}$$

The objective function (9.1) minimizes the sum of distributors' expected surpluses (and hence the firm's expected payments). Constraints (9.2) ensure cost coverage for all distributors. The first term on the left-hand side of this equation represents the expected weekly compensation that distributor k receives. Since the surplus variable s_k must be nonnegative (9.6), equation (9.2) ensures that the total expected compensation exceeds the distributor's anticipated cost C_k. Constraints (9.3) and (9.4) specify that fee values must increase with weight and distance, and constraints (9.5) impose the upper and lower bounds on fee values.

Formulation [FSP] is simple, yet remarkably versatile. First, although delivery quantities are stochastic, the model captures the impact of this stochasticity on distributor compensation, i.e., it accounts for the variability in stores' orders in computing the fees that distributors receive.

Second, the model can easily accommodate the alternative *minmax* objective of minimizing the maximum surplus over all distributors. Third, the model can also incorporate a variety of additional constraints to control the "shape" of the fee function (with respect to weight and distance). For instance, we can specify minimum and maximum permissible fee increments from one range to the next, and impose concavity constraints that ensure scale economies in weight or distance. Finally, we can easily adapt the model to represent alternative fee structures that specify incremental fees per mile or pound, rather than the fixed fee g_{ij}, in each cell of the table. Balakrishnan et al. (2000) discuss these enhancements in detail.

2.3 Model application and impact

The fee-setting linear program [FSP] is easy to implement and solve using standard linear programming packages on a personal computer. As we mentioned in Section 1, the manufacturing firm had recently upgraded its information system to integrate both internal and external operations (e.g., order processing, point-of-sale and distribution transactions). Consequently, sales and shipment data for each store were relatively easy to obtain. The firm generated estimates of distributors' total costs based upon the results of an activity-based model as well as detailed financial data provided by its distributors as part of their partnership arrangements. Bounds on fee values and increments were determined from representative commercial freight rates.

Based on these inputs, the model generated a new fee table that not only provided more equitable compensation to distributors but also reduced the firm's total expected payments compared to the option of using the firm's current (manually chosen) table.[1] In particular, using the current table, expected distributor surpluses ranged from 0 to 59% above cost. The optimized table (using the current ranges) ensured better parity among distributors by reducing the maximum surplus to under 28%, and simultaneously decreasing the firm's total expected payments by over 12%. Iterative application of the [FSP] model for different choices of weight and distance ranges led to further improvements in the fee table. Balakrishnan et al. (2000) present a more detailed discussion and interpretation of these results. A year after the optimized fee table was introduced, the distributors were supportive of the new fee-setting

[1]Due to cost increases, the fee values in the current table had to be inflated to ensure that no distributor makes a loss.

approach, prompting the firm to apply the optimization-based approach again to revise fee values.

In addition to these tangible benefits, the model provided a structured framework for managers to think about fee setting. Modeling constructs such as the cost coverage and fee credibility constraints facilitated managers' understanding of the structure of the fee-setting problem, its economic impact, and the complicated interrelationships among fee values. The [FSP] model shows how to obtain an accurate representation of distributor compensation by explicitly incorporating demand variability, and also provides diagnostic capabilities. By comparing distributor surpluses, planners can identify "inefficient" distributors whose operations and costs necessitate higher fees than do other distributors. Such analyses might suggest operational improvement initiatives, or perhaps even a reassignment of stores to distributors. The model also greatly facilitates sensitivity analysis. For instance, it permits studying the behavior of compensation as a function of variability in store demands. This type of analysis revealed that, under some circumstances, total compensation can actually decrease as demand variability increases. This counter-intuitive behavior occurs because, as demand variability increases, a higher proportion of demand might fall in lower weight ranges; if the fee values for these lower ranges are significantly smaller, then total fees paid to a distributor will decrease.

In summary, the fee-setting model proved to be an effective decision tool to support compensation negotiations with distributors. This model, which has not been previously addressed in the literature, is applicable to a variety of partnership settings where a uniform compensation scheme, based upon fee tables, is needed. The model also opens avenues for further research on compensation design. The [FSP] model focuses on fee setting, assuming that the weight and distance ranges are pre-specified. At a higher level, one can consider a model that simultaneously incorporates both range selection and fee setting decisions. Although easy to formulate, this problem is non-linear and therefore much more difficult to solve. Developing effective solution methods for this problem is a promising direction for further research.

3. Distribution Planning for Scheduled Deliveries with Uncertain Demand

The estimates of each distributor's expected weekly cost to meet delivery obligations serve as critical inputs for the fee-setting model described in Section 2. The manufacturer would like to accurately estimate these costs, both to ensure that distributors' operations are efficient and to reduce the overall fee payments needed to cover costs. In turn, these costs depend on the distributors' resource levels and delivery activities. Specifically, distribution costs consist of fixed costs for delivery resources (e.g., trucks and drivers) plus activity-dependent costs (e.g., costs that depend on trip mileage and duration). In addition, since big-box retailers' order sizes vary from week-to-week, deliveries to stores might entail extraordinary costs to accommodate unusually high demands. Consequently, accurate cost estimation requires modeling each distributor's operations. In this section we describe a distribution planning model that can support distributors' tactical decisions (e.g., acquisition and deployment of delivery resources), and can also serve as a cost estimation tool for the manufacturer. Given the stores served by each distributor and their projected demands, the cost-minimizing model incorporates resource-sizing decisions (i.e., choosing the appropriate truck fleet size), and determines a set of truck routes for each day of the week in order to meet all stores' demands. The manufacturer and its distributors can use the model to provide cost and efficiency benchmarks, assess the impacts of demand variability, and identify opportunities for improving distribution operations.

This section focuses largely on motivating and developing the distribution planning model, and only briefly discusses our implementation of a particular heuristic solution approach (based on genetic algorithm search techniques). Three characteristics of our distribution planning problem differentiate it from existing deterministic and stochastic vehicle routing problems in the logistics literature: (i) store deliveries must follow a fixed schedule; (ii) when demand is high, the distributor can use contingency resources to make deliveries, i.e., demand can "overflow" or exceed truck capacity, but at a cost; and (iii) since the same truck can serve different routes on different days, the model must incorporate "delivery calendaring," i.e., assigning truck-routes to days of the week. Section 3.1 reviews the big-box retailers' delivery requirements, and discusses their implications for efficient delivery strategies. In Section 3.2, we formally define and formulate the distribution planning problem as a stochastic optimization model. Section 3.3 briefly discusses an embedded overflow optimization subproblem needed to compute the cost coefficients for the model, and Section 3.4 outlines a genetic algorithm that we developed to approximately solve this problem. Section 3.5 reports preliminary computational results.

3.1 Distribution challenges and delivery strategies

Distributors face several challenges in meeting the big-box retailers' stringent supply requirements to support their high-volume, high-variety sales strategy. Prom the delivery planning perspective, the following three requirements are important:

1 *Scheduled deliveries:* each store requires periodic (typically weekly) deliveries, on the same day(s) and time each week.

2 *Unconstrained order quantities:* stores can order items in any desired quantities, without limits on minimum or maximum order sizes; orders can be placed as late as the day before a scheduled delivery.

3 *Single, consolidated delivery:* orders placed before each scheduled delivery must be delivered fully on a single truck–unless the order size exceeds a full-truckload.

Stores require single, scheduled deliveries due to the bottlenecks they face at their unloading docks. Since stores stock a wide variety of products supplied by numerous manufacturers, and since they require frequent deliveries (in small lot sizes) to minimize inventories, their loading docks are highly congested. Consequently, they require tightly coordinated deliveries from suppliers.

To illustrate the scope of distributors' operations, consider the requirements facing one distributor whose operations we studied in detail. This distributor supplied 67 stores within a 325-mile radius from the warehouse. The average distance from warehouse to store was 130 miles. Over 95% of these stores required one delivery per week; a few larger stores required multiple deliveries per week, while three small stores ordered products only once every two weeks. On average, stores ordered approximately 5,400 pounds[2], or one-eighth of a truckload, per delivery. But, order quantities varied widely from week-to-week. The coefficient of variation of demand (defined as the standard deviation of order weight divided by mean order weight), averaged across all 67 stores, was approximately 0.85; the ratio of maximum to mean order quantity ranged as high as 6.3 for a single store.

Let us now examine the implications of these delivery requirements in terms of viable strategies for the distributor. Since stores often order far

[2]Since the products are quite dense, truck capacities are constrained by the total weight rather than volume of products.

less than a full truckload and since they can be quite distant from the warehouse, visiting just one store on each trip is not cost effective. Instead, the distributor operates "milk runs" or multi-stop delivery routes, originating at the warehouse. We refer to the subset of stores served on a route under "normal" circumstances (i.e., when the total actual demand of these stores does not exceed the truck's capacity) as a *group*. Since each store requires delivery at the same time each week, the store groups are static. That is, the set of stores to be served on a route does not change from week to week.

If the weekly order quantity for each store were constant, then a deterministic vehicle routing algorithm (incorporating truck capacity constraints) might prove adequate to decide optimal (cost minimizing) store groups and truck routes. Even in this scenario, however, the model should incorporate fleet sizing and scheduling decisions, i.e., assigning groups to days of the week, and assigning trucks to groups in order to minimize the number of trucks needed to meet all delivery needs. On the other hand, if stores did not insist on strict delivery times but instead permitted deliveries at any time during their scheduled delivery day, then the distributor could dynamically decide the store groups and store visitation sequences each day–after observing demands. In this case, the associated planning decision consists of deciding the number of trucks to acquire in order to meet the total daily delivery requirements; a newsvendor type model might be appropriate for this problem. The big-box distribution context, however, is more complex than either of these two scenarios because order sizes vary significantly from week to week and stores will accept deliveries only at the scheduled time.

Distributors might employ two broad strategies to cope with variability in stores' demands while providing scheduled delivery services–maintaining safety capacity or using contingent capacity. The *safety capacity* approach refers to planning for some buffer capacity in the truck by selecting store groups such that, on average, the total weekly demand for the group is less than the capacity of the truck. Note that this strategy reduces the capacity utilization of trucks. *Contingent capacity* refers to the option of opportunistically using external resources–short-term truck rentals and temporary drivers or third-party delivery services–to handle demand *overflow*. That is, if on a particular day the total realized demand for all stores in a group exceeds the capacity of the *regular* truck (the distributor-operated truck that is assigned to this group), then the distributor "offloads" some of the stores from the regular truck, and uses outside resources to deliver to these stores. Embedded in this strategy are decisions regarding which stores to offload for each combination of store demand realizations, and how to route the trucks to cover the actual stores assigned to them. Distributors incur higher costs when they offload deliveries to outside sources compared to their regular delivery costs, i.e., contingent capacity is more expensive than regular capacity. Note that the safety capacity and contingent capacity strategies are not mutually exclusive, i.e., the distributor might plan for some buffer capacity in trucks while simultaneously relying on outside resources for high-demand scenarios.

Having examined the delivery constraints facing the distributor, and the delivery strategies to meet these constraints, let us now examine the distributor's cost structure. Weekly distribution expenses consist of *regular* costs to own and operate the distributor's delivery resources, and *overflow* costs when demand exceeds truck capacities. Regular costs have two components: *fixed* costs including truck leasing expenses, insurance, driver salaries, and administrative overhead, and *activity-dependent* costs that depend on actual mileage and duration of delivery routes such as gasoline costs, mileage charges, and overtime costs. Route duration includes both driving time and the time needed to unload orders at each store, and so activity-dependent costs can also depend on the number of stores visited on a route. *Overflow* costs can include a fixed cost for each overflow occurrence as well as activity-dependent components. If overflows are handled by renting trucks temporarily, the fixed cost might represent the daily truck rental charge. On the other hand, if the distributor relies on a commercial package delivery service or a less-than-truckload carrier for overflow deliveries, then the overflow cost might be additive across stores, depending only on the delivery weight and distance from the warehouse to each individual store. Due to demand variability, only the fixed regular costs for a given distribution plan can be determined with certainty; activity-dependent costs and even the fixed portion of overflow costs will not be known until demands are realized. Since stochastic demands cause uncertainty in costs *a priori,* we minimize *expected* weekly costs during the planning phase.

To cost-effectively satisfy stores' service requirements, distributors must carefully plan their acquisition and deployment of resources. Regular fixed costs depend on resource acquisition or capacity planning decisions. Because activity-dependent routing costs and contingency costs are also significant (accounting for more than one-third of total delivery-related costs for the distributor we studied), the capacity-planning decisions should be made in conjunction with detailed routing considerations, and should reflect the impact of stochastic demands. When making these planning decisions, distributors face numerous complex tradeoffs. What is the best compromise between maintaining safety capacity and using contingent capacity? Should stores with relatively stable demands be grouped together so that these groups have minimal overflows, or should each group contain a mix of stable- and variable-demand stores? Should stores near the distribution center be grouped onto one route, or should such stores be dispersed across routes to serve as convenient stores to remove from routes when overflow occurs? To provide answers to such questions, we next propose a quantitative model that considers both resource *capacity planning* and resource *deployment.*

3.2 Problem definition and formulation

Given the set of stores served by the distributor, their respective locations, delivery frequencies, and demand (probability) distributions, we wish to minimize the distributor's expected weekly distribution expenses, including regular and overflow costs, to meet these demands. Costs depend on the following interrelated tactical decisions:

- *Resource (fleet and driver) sizing:* deciding the number of regular trucks to acquire, and number of full-time drivers to employ;
- *Store grouping:* forming groups of stores to be served by regular truck routes;
- *Truck assignment:* assigning each chosen group to an available (regular) truck; and,
- *Delivery calendaring:* deciding the day of the week on which each group will be served.

For simplicity, let us assume that each store requires one delivery per week.[3] Since stores require single, consolidated deliveries, we must assign each store to exactly one group. Store grouping decisions might be constrained in various ways. For instance, human resource policies might dictate that a regular tour cannot extend more than two days; in this case, each group must be chosen such that a truck can visit and drop off loads at all stores in the group and return to the warehouse within two days. Similarly, for operational reasons, the distributor might impose an upper limit on the number of stores in each group. The delivery calendaring decision, in conjunction with the truck assignment decisions, must ensure that trucks are only assigned to one group at a time. Again, this decision might be constrained if individual stores have requested deliveries on specified days of the week.[4]

[3]If a store requires multiple deliveries in a week, we can equivalently replicate the store so that each replicate orders once per week.

[4]Negotiating delivery day and time with stores might be an iterative process. Based on current plans, the distributor might propose the delivery day and time to each store. If the

The expected activity-dependent costs for regular trucks and expected overflow costs depend on several factors: store grouping and truck assignment decisions, variability in store demands, and the policies used for truck routing and overflow handling. These latter policies include:

- *offloading policy:* deciding which stores to offload from the regular truck when total realized demand for a group exceeds the assigned regular truck's capacity; and,

- *visitation sequencing policy:* deciding how to route (regular and possibly overflow) trucks once stores have been offloaded.

We will later elaborate on the cost tradeoffs that these policies must consider. For brevity, we will refer to the activity-dependent costs for regular trucks as regular routing costs.

Our model assumes that the regular routing costs and overflow costs are separable by group. That is, for a particular group, these costs depend only on the stores assigned to that group (and the associated truck assignment, as well as the offloading and visitation policies), independent of store assignments to other groups. The offloading policies that distributors use in practice support this assumption. For instance, separability of regular routing costs holds because tight delivery-scheduling constraints preclude assigning an offloaded store from one group to another group (for which the regular truck might have available capacity) scheduled for the same day. Similarly, overflow costs are separable if distributors do not combine offloaded stores from different groups into a single overflow route, or if the cost of overflow deliveries to individual stores are additive.

We next present a mathematical formulation of the *distribution planning problem with stochastic demands and scheduled deliveries,* abbreviated as the *DPSS* problem. Let $i = 1, \ldots, n$ index the n stores that a given distributor must serve. Trucks can be of different types, with varying capacities and different fixed and operating costs. For each truck type, suppose we can determine the maximum number of regular trucks needed of that type, using, for instance, rough-cut overall demand-capacity analysis or industry norms. Let K denote the total number of candidate regular trucks (over all truck types) needed to serve the n stores. We index the candidate trucks from $k = 1, \ldots, K$, and associate a truck type with each truck index k. Let F_k denote the fixed cost (per week) for truck k, and let B_k be truck k's capacity. Suppose

store cannot accommodate this schedule, the distribution plan needs to be revised so that the group containing that store is assigned to the required day.

we enumerate all feasible groups of stores, and index these groups from $g = 1, \dots, m$. A group of stores is feasible if it meets restrictions on tour duration and group size, as well as other requirements (e.g., we might require that two proximate stores must always be assigned to the same group, or two distant stores should never be assigned to the same group). For $i = 1, \dots, n$, let $G(i)$ be the set of all groups g that include store i.

Let C_{gk} and D_{gk} denote, respectively, the *expected* regular routing cost and *expected* overflow cost to serve all stores in group g if regular truck k is assigned to this group. We defer discussion on how to compute these costs in order to first present the overall DPSS formulation. To simplify the exposition, we assume that a truck can be assigned to at most one group on any day of the week. If T denotes the number of days of the week when stores accept deliveries (e.g., $T = 5$ or 7), let $t = 1, \dots, T$ index the available delivery days in the week.

To model the various decisions of the DPSS problem, we define the following binary decision variables:

Group selection:	$y_g = 1$,	*if we select group g, and 0 otherwise, for all* $g = 1, \dots, G$;
Group-truck-day assignment:	$x_{gkt} = 1$,	*if we assign group g to truck k for deliveries on day t, and 0 otherwise, for all* $g = 1, \dots, G$, $k = 1, \dots, K$, *and* $t = 1, \dots, T$; *and,*
Truck selection:	$z_k = 1$,	*if we acquire truck k, and 0 otherwise, for all* $k = 1, \dots, K$.

Using these decision variables, we can formulate the DPSS problem as follows:

[DPSS]

$$minimize \sum_{k=1}^{K} F_k z_k + \sum_{g=1}^{m} \sum_{k=1}^{K} (C_{gk} + D_{gk}) \sum_{t=1}^{T} x_{gkt} \qquad (9.7)$$

subject to:

$$\text{Store assignment}: \quad \sum_{g \in G(i)} y_g = 1, \qquad i = 1, \ldots, n, \tag{9.8}$$

$$\text{Group assignment}: \quad \sum_{k=1}^{K} \sum_{t=1}^{T} x_{gkt} = y_g, \qquad g = 1, \ldots, m, \tag{9.9}$$

$$\text{Truck assignment}: \quad \sum_{g=1}^{m} x_{gkt} \leq z_k, \qquad k = 1, \ldots, K, \tag{9.10}$$

$$t = 1, \ldots, T, \; and$$

$$\text{Integrality}: \quad x_{gkt}, y_g, z_k \in \{0, 1\}, \quad for\ all\ g, k, \; and\ t. \tag{9.11}$$

The objective function (9.7) minimizes the total fixed and activity-dependent regular costs and overflow costs. The store assignment constraints (9.8) specify that each store must be assigned to exactly one selected group. Constraints (9.9) relate the group selection decisions to the assignment of groups to truck-day combinations. These constraints specify that if we select a group g (i.e., if $y_g = 1$), then this group must be assigned to exactly one truck-day combination. Note that we can incorporate store delivery-day preferences by including only the relevant assignment variables, thus reducing the size of the formulation. That is, if a store in group g cannot accept deliveries on day t', then we can omit the variable $x_{gkt'}$ for all truck indices k. Constraints (9.10) serve both as truck-scheduling restrictions and truck-selection forcing constraints. We can assign a truck k to any group and day only if we select the truck (i.e., if $z_k = 1$); and, if we select the truck, the constraint specifies that we cannot assign it to more than one group on the same day.

The DPSS model differs from classical deterministic and stochastic vehicle routing models in several respects; the model considers stochastic demands, overflow options, truck-to-route assignments, and delivery calendaring (i.e., route-to-day assignments). Deterministic vehicle routing models, for instance, ignore variability in demands, and impose truck capacities as hard constraints. Stochastic vehicle routing models typically do not incorporate shared resources (trucks) across routes, and ignore the delivery calendaring problem (see Bertsimas and Simchi-Levi, 1996, Gendreau, Laporte, and Seguin, 1996, and Stewart and Golden, 1983, for a discussion of the scope of stochastic vehicle routing models found in the literature).

The distribution planning problem is a large-scale stochastic integer programming problem with recourse (see Birge and Louveaux, 1997, Higle and Sen, 1996). Under the stochastic programming framework, decisions regarding which trucks to choose and how to group the stores constitute the first stage decisions. The recourse decisions entail determining, for each possible realization of store demands, which stores to offload and how to route regular and overflow trucks to minimize the regular (routing) and overflow costs. Formulation [*DPSS*] explicitly represents the first stage decisions of group and truck selection and assignment. The cost coefficients C_{gk} and D_{gk} reflect the expected costs of the second stage overflow offloading and (regular and overflow) truck routing decisions. Stochastic integer programming problems are notoriously difficult to solve. Note that the number of feasible groups, m, in formulation [*DPSS*] can be exponential in the number of stores. The structure of formulation [*DPSS*] suggests using a column generation approach to solve the problem or its linear relaxation. As we discuss next, the costcoefficients C_{gk} and D_{gk} for each group can be difficult to compute since they entail solving subproblems that are equivalent to or more difficult than the traveling salesman problem.

3.3 Determining the expected routing and overflow costs

Our cost separability assumption implies that we can compute regular routing costs and overflow costs independently for each group g. Given the set of stores in a group, the corresponding regular and overflow cost coefficients C_{gk} and D_{gk} depend on several factors–the underlying cost drivers (e.g., delivery weights and driving distances), overflow handling policies, and visitation sequencing. Estimating the cost coefficients, therefore, entails solving an embedded optimization problem in order to select routing and overflow handling policies that minimize $C_{gk} + D_{gk}$ for each store group g and assigned truck k. We refer to this embedded problem as the *overflow optimization subproblem*[5], and note that it is a stochastic optimization problem because it must consider the random variations in store demands.

Let us now consider some overflow modeling assumptions and options that apply to the big-box distribution context. In this setting, tight delivery schedules and large distances between the warehouse and certain stores prohibit regular trucks from handling overflow demand. So, when an overflow occurs, the routing cost for the regular truck is proportional to the travel distance (and time) to serve only those stores that have not been offloaded. In contrast, Gendreau, Laporte, and Seguin (1996) consider a vehicle routing problem with stochastic demands in which, if the total demand for all stores in a group exceeds the truck capacity, the regular truck returns to the warehouse for reloading at

[5]Note that, although we refer to the subproblem as the "overflow" optimization problem, it also determines the activity-dependent components, C_{gk}, of regular costs.

appropriate intermediate points in the route. Gendreau et al. assume that the (regular) truck visits the stores in the same sequence regardless of the actual demand realization. We make a similar *static visitation sequence* assumption for the big-box distribution context. Our overflow optimization model decides the relative order in which stores must be visited. When an overflow occurs and stores are selected for regular and overflow deliveries, the truck (regular or overflow) follows the specified visitation sequence but only visits those stores to which it must deliver. This assumption is appropriate for big-box deliveries since there is little opportunity to dynamically vary the visitation sequence because stores are not willing to accept deliveries before or after their scheduled time.

Two other problem characteristics determine the structure of the overflow optimization model–the offloading policy, and the overflow cost structure. We next discuss the choices along each of these dimensions.

- *Offloading policy:* To decide which stores to offload when an overflow occurs, we might use either a *static offloading policy* or a *dynamic offloading policy.* Under a static policy, the sequence in which stores from the group must be offloaded is predetermined (by the overflow optimization model); when an overflow occurs, stores are offloaded in this specified sequence until the demand for the remaining stores is less than or equal to the regular truck's capacity. In contrast, dynamic offloading policies make state-dependent offloading decisions based on the actual store demand realizations. Dynamic offloading policies are likely to be more cost-effective; but, they vastly complicate the overflow optimization problem, and can be difficult to implement.

- *Overflow cost drivers:* Overflow costs might be either *route-based* or *store-based* depending on the outside delivery service option that the distributor uses for overflow deliveries. With route-based costs, the overflow delivery cost depends on the actual distance that the truck travels and/or the actual duration (including unloading time at stores) of the route. In contrast, with a store-based cost structure, the delivery cost decomposes by store, i.e., the cost to deliver to a subset of stores is the sum of costs for delivering to each store. This store delivery cost might depend on the weight of items delivered and the store-to-warehouse distance–similar to the tariff structure that commercial carriers use for less-than-truckload shipments. Note that store-based cost structures simplify the overflow optimization problem, making overflow delivery sequencing decisions irrelevant.

For our tactical planning model, we will assume that the distributor uses a static offloading policy. In this class of policies, we highlight one particularly interesting offloading rule that relates the offloading sequence to the visitation sequence. This rule, which we call the *reverse visitation* rule, offloads stores in reverse visitation order. That is, when an overflow occurs, stores are offloaded in last-visited-first-offloaded order until the remaining demand fits in the regular truck. Note that, by returning to the warehouse for reloading at intermediate points in the fixed visitation sequence, Gendreau et al.'s model implicitly follows this offloading rule. The reverse visitation rule is easy to implement in practice, and because this policy assigns "contiguous" stores from the original visitation sequence to both the regular and overflow trucks, it avoids additional intermediate truck idle time between scheduled store deliveries. Moreover, making the offloading rule a function of the visitation sequence eliminates one set of decisions, thus simplifying the overflow optimization problem.

The overflow optimization problem entails deciding the static visitation and offloading sequences that minimize the expected routing plus overflow costs to service the stores in the group. We do not present a formal mathematical formulation for this problem, but instead examine one of its ingredients, namely, how to compute the expected costs for a given (static) visitation sequence and offloading sequence (not necessarily the reverse visitation sequence). Let L denote the number of stores in the group. Suppose we index the stores from 1 to L in the order in which they will be offloaded if we follow the given offloading sequence, i.e., in the event of an overflow, store 1 is offloaded first, and so on. For $l = 0, \ldots, L$, let $R_l = \{l+1, \ldots .L\}$ denote the set of stores that the regular truck services when l stores are offloaded; let $O_l = \{1, \ldots, l\}$ be the complement of R_l, i.e., the set of offloaded stores. For expositional convenience, assume either that the overflow costs are store-based or the overflow truck has unlimited capacity (so that at most one overflow truck is needed for any group on a given day). For $l = 0, \ldots, L$, let c_{lk} and d_l denote, respectively, the expected cost for the regular truck k and the expected overflow cost if l stores are offloaded. The regular cost c_{lk} is easy to compute since the visitation sequence is given (otherwise, we must solve a traveling salesman problem over the set of locations R_l). Similarly, if overflow costs are route-based, the given route determines the value of d_l. If overflow costs are store-based, then d_l is the sum of expected delivery costs to each store in O_l, which depends only on the warehouse-to-store distance and the probability distribution of demand for each store. Now, let p_{lk} denote the probability that exactly l stores will be offloaded (i.e., the probability that the total demand for all the stores in the set R_l is less than or equal to the truck capacity B_k, *and* adding the demand of the l^{th} offloaded store exceeds the truck capacity). Then, the expected regular and overflow costs corresponding to the given visitation sequence and offloading sequence are:

$$C_{gk} = \sum_{l=0}^{L} p_l c_{lk} \ and \ D_{gk} = \sum_{l=0}^{L} p_{lk} d_l. \tag{9.12}$$

The overflow optimization subproblem requires deciding the visitation and offloading sequence that minimizes the sum of these two expected costs. In our solution procedure, described in the next section, we do not solve the overflow subproblem optimally, but instead use heuristic rules to decide visitation and offload sequencing.

3.4 Genetic algorithm for the DPSS problem

As our discussions in Section 3.3 suggest, solving the DPSS problem optimally is practically impossible. The problem has a vast number of decision alternatives (e.g., store groupings, group-truck-day assignments), and even computing the expected cost for a given store group and truck assignment can be intractable. We, therefore, developed a genetic algorithm (GA) (see, for example, Goldberg 1989) that iteratively generates distribution plans (store groupings and group-truck-day assignments). We provide only a sketch of the algorithm here, omitting several details of our implementation.

Since the number of regular trucks needed varies only within a relatively small range, our GA does not incorporate fleet sizing decisions, but takes as input the desired number of trucks. By changing the fleet size and reapplying the GA, we can determine the number of trucks needed to achieve the lowest-cost solution. The central algorithmic decision when applying GA to any optimization problem concerns the representation of feasible solutions as genes. Let K denote the desired number of trucks, and let T be the number of delivery days in a week. For expositional convenience, we assume that each truck can handle at most one multi-stop route each day; some routes might require multiple days. Instead of considering group-truck combinations as in formulation $[DPSS]$, we now consider truck-day combinations. We index the truck-day combinations from $j = 1$ to KT, starting with the first day for the first truck; thus, the index j corresponds to using truck number $k = \lfloor j/T \rfloor$ on day $t = j \bmod (T)$. Note that the number of truck-day combinations, KT, also serves as an upper bound on the number of groups that will be chosen. If n denotes the total number of stores served by the distributor, we represent the gene as an integer vector V of length n. The i^{th} position of this vector (the i^{th} allele) can take any integer value $v(i)$ between 1 and *KT;* this value represents the index of the truck-day combination assigned to store i. Thus, all stores assigned to the same truck-day combination constitute a group. If the full route to cover all of the stores in a group assigned to truck k on day t takes multiple days, then day t represents the *starting* day for the route.

The value or fitness function of a gene specifies the total regular and overflow costs for all of the grouping, truck assignment, and scheduling choices implied by the alleles of the gene. Thus, evaluating the fitness of each gene requires solving the overflow optimization problem for each group-truck assignment that the gene selects. The value also includes penalties for violating constraints on group sizes and tour durations. For instance, if the distributor specifies that a group can contain no more than 10 stores, then any gene that assigns more stores to a group receives a high penalty. Similarly, we use penalties to preclude tours that require more than two days. We also assign a high penalty if one group requires a two-day route, and the truck assigned to this group is also assigned to start deliveries to another group on the second day.

Starting with an initial generation of genes, the GA generates successive generations through the standard neighborhood-defining operations of crossovers, mutation, and selection (see Goldberg, 1989). For the DPSS application, mutation corresponds to randomly changing the current value $v(i)$ of an allele in a particular gene to another value between 1 and *KT*. Crossovers correspond to interchanging a range of allele values between pairs of randomly chosen genes. Given a generation of genes, members in the next generation are selected based on the fitness values for the members and their neighbors from the previous generation. The method stops when the best solution does not improve over successive generations or upon reaching the maximum specified number of iterations.

3.5 GA implementation

We implemented the GA on a personal computer, using C++. Our implementation does not solve the overflow optimization problem optimally in the algorithm's fitness evaluation module. Instead, we incorporate heuristic rules to select the (static) visitation and offloading sequences, and apply a method similar to the approach used to derive the cost expressions in (9.12) to estimate the expected routing and overflow costs for each group-truck combination. Because the big-box stores tend to be co-located within major population centers, the topology of the distribution network resembles a tree network connecting dense store clusters. These store clusters are relatively easy to identify by imposing limits on the diameter of each cluster and visually examining maps of a distributor's coverage region. Given the group of stores to be covered on a route and the "cluster" to which each store belongs, we used a simple "sweep" heuristic to decide the sequence in which to visit the clusters. Approximations of inter-cluster and intra-cluster distances, together with estimates of average driving speeds, simplify route distance and duration calculations. We tested various alternative offloading policies–offloadinglargest demands first, offloading closest stores first, and offloading in reverse visitation order; based on preliminary results, we chose to use the reverse visitation rule. Note that, although we used approximate procedures, the GA framework permits using more sophisticated fitness evaluation routines, including exact methods to solve the overflow optimization subproblems.

Based on our analysis of weekly demand patterns, we assumed that store demands are normally distributed; the mean and variance of demand for each store were estimated from historical data. Our routing and overflow-costing model assumes that store demands are independent. When analyzing route overflow conditions, we permit using multiple (capacitated) overflow trucks per store group, up to a maximum number. For our data set, we found that the likelihood of requiring five or more overflow trucks was negligible, and therefore limited the number of overflow trucks per store group to five.

Based on preliminary computational experiments, we chose the required GA parameter settings such as crossover and mutation probabilities. We did not attempt to systematically compare the relative performance of competing GA approaches, or to fine-tune the performance of the algorithm. Our immediate goal was rather to simply apply a standard genetic algorithm for the purpose of answering the two primary questions that motivated the distribution planning exercise:

- can we accurately estimate distributors' delivery costs via a robust model-based approach, considering all the associated problem complexities, including stochastic demand?; and,
- does the extent of retail store-order variability significantly influence distribution costs?

3.6 Preliminary computational results

Our initial efforts during the model-testing phase focused on model validation. The manufacturer had previously developed planned routes using a deterministic route-optimization software package and, working with distributors, had derived cost estimates for each route using a detailed activity-based costing approach. For the distribution operation (covering 67 stores) that we studied, we first applied our fitness function evaluator to the manufacturer's planned routes. Inputs to the evaluator included actual store locations and underlying cost parameters (e.g., per-mile and per hour costs for regular and overflow deliveries, truck fixed costs) provided by the manufacturer. Assuming (as the manufacturer's cost estimation method did) that store demands are deterministic (and equal to their average demands), the costs predicted by our model for the store groups proposed by the manufacturer were remarkably close to the manufacturer's estimates–the costs differed by less than two percent. These results confirmed that our modeling approximations (e.g., store clustering and using the sweep heuristic for visitation sequencing) were appropriate.

Next, using the planned routes chosen by the manufacturer (derived via a deterministic algorithm), we introduced stochastic demands and applied the DPSS model to determine the expected total cost as a function of variability. Demand variability can be easily adjusted in the model by specifying alternative values for the demand *coefficient of variation* (*COV*) for all stores served by the distributor. Figure 9.1 shows how expected total cost varies as demand becomes increasingly uncertain; distribution costs are relatively low when variability is low, but increase with variability. When COV increases from 0 to 1, for example, total cost increases approximately 20%. Although the building-products manufacturer had previously recognized that store-order variability resulted in high costs for the distributors, the firm found our DPSS model particularly useful for quantifying the cost of that variability.

Finally, we used the GA to generate new near-optimal store groupings and truck-day assignments. We wished to compare the expected costs of our solutions to the manufacturer's planned routes and also study the impact of store-order variability on delivery costs–providing the manufacturer with results that it could not previously obtain. Figure 9.1 compares the cost performance of GA-optimized routes with the manufacturer's proposed solution. The figure shows results for two choices of regular trucks–3 trucks and 4 trucks (the planned routes used by the manufacturer assumed 4 trucks for this distributor); for each choice, the figure shows the total cost of the best GA solution at each value of COV. As these preliminary results suggest, genetic algorithm techniques can serve as a powerful method for generating good tactical plans for distributing products to big-box stores.

Our computational tests also showed that distribution plans optimized for low levels of variability tended to perform poorly when uncertainty increased, whereas plans designed for uncertain conditions were more

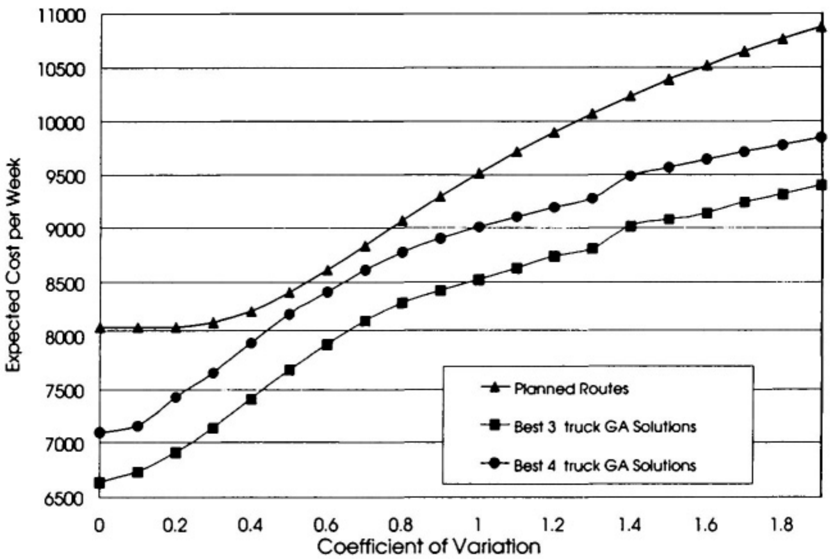


Figure 9.1. Expected cost of planned routes versus store-order variability

robust to changing variability. This finding implies that routes derived from deterministic routing packages might prove very costly in practice because such packages strive for high truck utilization and ignore the possible overflow consequences.

4. Controlling Variability in the Distribution System

The two problems considered in Sections 2 and 3 stressed the need for modeling approaches that account for the impact of variability in store orders on distributor compensation and delivery costs. Suppliers facing high demand variability also require significant investments in inventory to meet high customer service level expectations. The building-products supply chain context we studied provides an excellent example of the burden order variability places on the system. Due to large uncertainties in stores' order quantities, distributors often experience either:

- insufficient stock and/or transportation capacity to meet the unusually large orders, or

- significant amounts of excess stock and idle transportation capacity due to unexpectedly small orders.

The short delivery lead times and high fill rates required by big-box retailers exacerbate these problems, forcing distributors to make last-minute arrangements for additional transportation capacity or expedited supply shipments at significant added cost.

The majority of supply chain inventory models discussed in the literature account for exogenous demand variability when optimizing strategic and operational decisions at each stage of the chain (e.g., Clark and Scarf, 1960). But, it is equally important to assess the extent to which actions and policies of channel members internal to the supply chain unnecessarily contribute to increased upstream demand variability, and to identify steps to decrease this variability in order to reduce system-wide safety stock and transportation costs. In this section, we expand the scope of our analysis of distributor-retailer interactions, encompassing both inventory and transportation costs, in order to assess the impact of demand variability at different stages in a supply chain. However, unlike the distribution planning model of Section 3, we take a more aggregate view of distribution operations, ignoring the detailed truck routing and scheduling issues.

To study the impact of store ordering policies on system costs, we consider order variability along two dimensions–SKU variability and load variability. *SKU variability* refers to week-to-week variation in the total order amount (measured, say, in pounds) for a particular stock-keeping unit (SKU) by *all* stores served by a distributor. Clearly, this variability affects the distributor's safety stock costs for that SKU since safety stock is typically proportional to the standard deviation of total weekly demand. On the other hand, as we have seen in Section 3, transportation costs depend on each store's *load variability,* the week-to-week variation in the total weight of all items that the store orders. Note that since SKU variability requires aggregation of demand for a particular SKU over stores, whereas load variability aggregates demand over items for a particular store, it is possible to have high SKU variability for some products and still have low load variability (e.g., if a store's orders for different SKUs are negatively correlated).

Our analysis of variability-propagation in the distribution chain begins with an examination of transactional data history to assess the scope of the problem in the building-products distribution network. Section 4.1 reports the results of our diagnostic analysis, confirming the cost burden that stores' variable ordering policies place on distributors. We then describe, in Section 4.2, a stylized distribution system model representing the distributor-retailer system, and present variance-damping mechanisms intended to increase channel coordination by reducing upstream (i.e., store order) variance in the supply chain. Finally, based on our computational test results, we assess the potential system-wide savings from using one of our proposed variance reducing policies.

4.1 Variation propagation in multi-stage systems: The bullwhip effect

Recent literature has emphasized the so-called "bullwhip effect" in supply chains–the tendency for demand variability to increase at upstream stages in the supply chain. Lee et al. (1997) cite several factors causing the bullwhip effect under rational decision making on the part of channel members, and suggest methods (such as information sharing and strategic partnerships) to decrease the amount of variance amplification in the supply chain (see Simchi-Levi et al. 1999). To assess the extent to which the bullwhip effect exists in the big-box distribution chain, we first used the available supply chain transactional data to compare week-to-week variability in end-customer demand (faced by the stores) with the current variability in stores' order quantities. The data consists of detailed POS (point-of-sale) transactions at each store served by one distributor, orders from stores to the distributor, and actual shipments from the distributor to stores. Of the 67 stores served by the distributor during the time frame for our analysis, 62 stores were open for more than one year; so, we consider only these 62 stores in our analysis of variability.

Let us first compare the load variability in end-customer demand with the load variability in store orders. Henceforth, we refer to end-customer demand facing the stores as "demand," and stores' orders to the distributors as "orders;" all quantities are measured in pounds of product. For two selected stores, Figure 9.2 shows the total demand, aggregated over all SKUs, and total orders received by the distributor from each store in each week of a 26-week period. In this figure, we have divided the actual demand and order quantities by their respective mean values over the 26 weeks, and show only the deviation from the mean (i.e., the y-axis corresponds to the ratio of each week's demand or orders to its mean). As the charts show, while the demand facing each store was relatively stable over time for both stores, the orders these stores placed to the distributor exhibited much greater variability, especially in the case of Store A.

If σ_o and σ_d denote a store's standard deviation of weekly orders and weekly demand, respectively, then the ratio σ_o/σ_d measures the store's *load variability amplification.* For Store B, σ_o/σ_d equals 2.52 for the 26-week period shown in the graph. Store A, with $\sigma_o/\sigma_d = 5.94$, amplifies its demand significantly more.

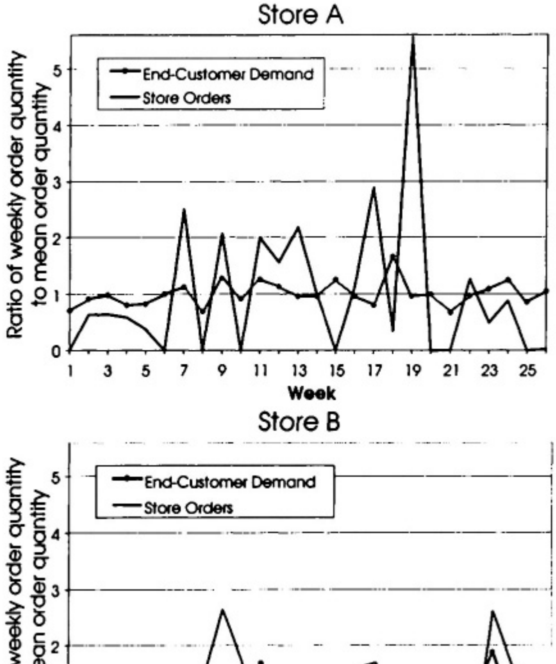


Figure 9.2. Illustration of demand variability amplification for selected stores

Turning to SKU variability, Figure 9.3 shows, for two representative SKUs, the total demand at all 62 stores (relative to mean demand) and total quantity ordered from the distributor in each week over a 26-week period. The amplification of demand variability (measured by the ratio of the ν_o to ν_d, where ν_o and ν_d denote respectively the standard deviation of weekly SKU orders and weekly SKU demand for all stores) equals approximately 1.5 for SKU 1 and 1.8 for SKU 2. Although amplification in SKU variability is not as severe as the amplification of load variability, these ratios still represent a 50% increase in the standard deviation of orders seen by the distributor for SKU 1 and an 80% increase for SKU 2, resulting in a costly safety stock burden at the distributor.

Similar analysis of load and SKU variability for other stores and SKUs showed that store ordering practices contributed significantly to variability amplification in the upstream stages of the distribution system, leading to both disruptions in weekly delivery planning (or increased transportation capacity requirements) and increased safety stock levels.

4.2 Upstream variability control mechanisms

With this evidence of variability amplification by stores, the manufacturer was interested in exploring alternate store inventory management policies that might reduce variability and wanted to estimate and the economic impact such reduction might have on logistics costs. Recent research has focused on coordinating supply chain operations through various mechanisms such as quantity- and time-flexible contracts (e.g., Tsay, 1999, Li and Kouvelis, 1999, Moinzadeh and Nahmias, 2000), synchronized and extended ordering cycles for multiple retailers (e.g, Cachon, 1999), linear transfer payments (e.g., Cachon and Zipkin, 1999), and innovative accounting systems that align individual and system-wide incentives (e.g., Lee and Whang, 1999, Chen, 1999). Baganha and Cohen (1998) considered the relationship between demand variability and inventory costs at various echelons in a supply chain, and presented conditions under which variability should be dampened to promote better system performance. Our approach considers the impacts of demand variability on both inventory and transportation costs in the supply chain, and develops control mechanisms to mitigate variability between supply-chain stages. We next introduce and motivate these variability-damping mechanisms, and demonstrate their effectiveness via numerical studies based on actual data from the building-products manufacturer.

Each variability-damping mechanism we propose provides an ability to *tune* the level of store-order variability by changing some specific system or policy parameter. Figure 9.4 provides a conceptual representation of

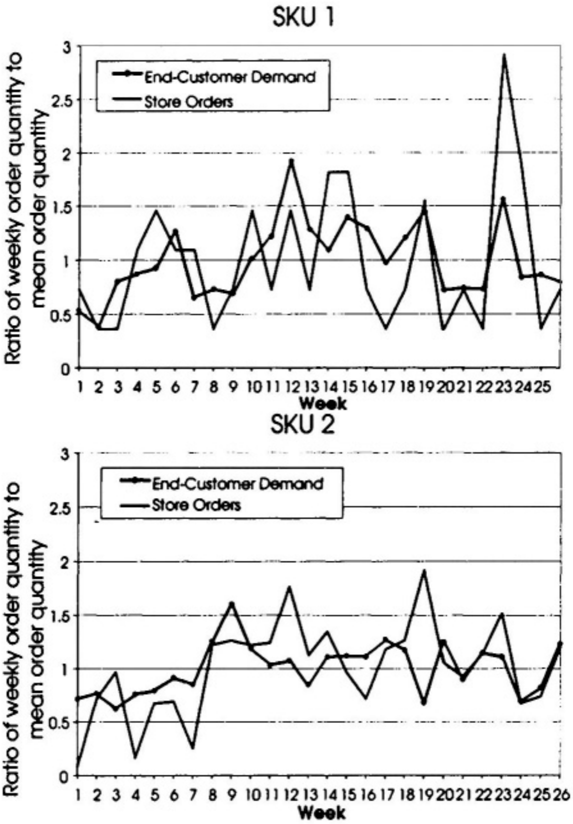


Figure 9.3. SKU demand variability amplification for selected SKUs

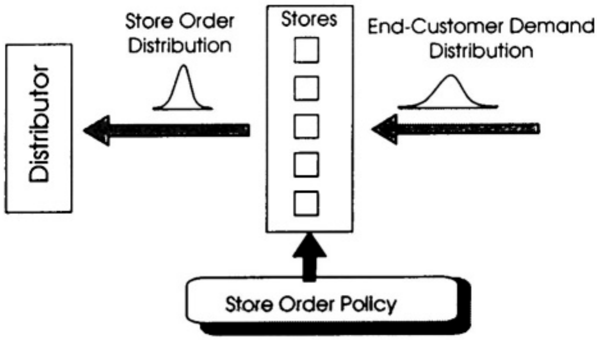


Figure 9.4. Store ordering policies as upstream variance control mechanisms

how a store's ordering policy serves as a control mechanism to influence the variability of orders received by the distributor. By constructing a model of the distributor-retailer distribution system and expressing system cost as a function of a specific policy parameter, we seek the best parameter value, and hence variability level from a system-wide view. Our approach effectively treats the variability of flows at the distributor-retailer interface as decision variables. We next summarize the important elements of our modeling approach (for more details on the model see Geunes, 1999).

4.2.1 Distributor-retailer modeling approach. Our model focuses on the operations of a single distributor supplying its assigned stores. Our goal is to develop an expression for *total system-wide cost* (including both the distributor's and stores' inventory and logistics costs) for a given class of store inventory management (and ordering) policies, assuming that all stores follow the same policy type, but with policy parameters that are tuned to their end-customer demand. The purpose of the model is to capture the relevant dynamics and costs of the distributor-retailer portion of the building-products supply chain and to demonstrate the potential for variance-damping mechanisms to reduce supply chain costs. We consider the distributor's inventory and transportation costs, as well as stores' inventory costs. We assume that stores can backorder retail demand if they run out of stock during a replenishment interval, and distributors can expedite supplies from the manufacturer in order to fully meet stores' orders each week. So, our system-wide cost includes the costs associated with these backlogging and expediting options. Since both the stores and distributor order periodically (and the ordering period is predetermined), we do not consider fixed setup or ordering costs.

Distributor transportation costs

As we noted in Section 2, retail stores served by the distributor are generally clustered together into one of several metropolitan areas. Our methodology creates a separate transportation cost model for each area or cluster[6] of stores after aggregating the demand for all stores in that cluster. As in Section 3, distributor transportation cost components include:

- *Regular fleet and driver costs:* cost of leasing and operating a fleet of *regular* trucks (including the costs for corresponding drivers) for weekly deliveries to cluster; and

- *Overflow costs:* cost of extra shipments when demand for a cluster exceeds the total capacity of regular trucks assigned to that cluster.

Instead of the detailed model of overflow management employed in Section 3, we now assume the overflow cost is linear in the total excess or overflow weight, as would typically be the case when the overflows are carried within a local region by a third party LTL carrier. We also do not consider detailed routing and truck assignment decisions. Let C denote the number of store clusters in the distribution region. For $c = 1, \ldots, C$, the estimated weekly driver, mileage, and truck lease costs combined with the expected distance and time for shipments to cluster c provide an estimate of the (weekly) fixed cost for each truckload shipment to the cluster, which we denote by K_c. Our model treats each store cluster as a single aggregate store, and then determines the regular truck capacity to be allocated to the cluster (in integer multiples of full truckloads). It then computes the expected transportation cost for each cluster based on the cluster's order distribution, allocated regular transportation capacity, and fixed plus volume-dependent shipping costs. Let K_{rc} and K_{oc} denote the cost per unit weight shipped via regular and overflow capacity to cluster c, and let T_c denote the transportation capacity allocated to cluster c. If x^c is the random variable for the total weekly ordered weight in cluster c, and $f(x^c)$ denotes the probability density function (pdf) of x^c, then the expected weekly transportation cost for

[6]Here "cluster" refers to stores within the same geographical proximity, not the store groups discussed in Section 3.

cluster c, $G(T_c)$, is given by

$$G(T_c) = (K_c + K_{rc})T_c - K_{rc}\int_0^{T_c}(T_c - x^c)f(x^c)dx^c + K_{oc}\int_{T_c}^{\infty}(x^c - T_c)f(x^c)dx^c. \quad (9.13)$$

We can show that $G(T_c)$ is convex in T_c when $K_{oc} \geq K_{rc}$, i.e., when the cost of overflow capacity is greater than or equal to the cost of regular capacity. The convexity of Equation (9.13) allows us to easily determine the optimal value of T_c (Equation 9.13 is similar to the classical newsvendor equation in single-period inventory analysis; see Nahmias, 2000).

Distributor inventory and shortage costs

We now consider the distributor's inventory-related costs. Let j index the set of products stocked by the distributor. The distributor attempts to fill all orders placed in a period from its own stock; if stock on hand is not sufficient, the distributor obtains expedited shipments from the manufacturer (who is assumed to have unlimited capacity) in time to meet all store orders. Expediting product j entails an additional expediting cost of e_{dj} per unit of product. The distributor incurs a holding cost, h_{dj}, per unit of inventory of product j remaining at the end of a period, and incurs a cost of c_{dj} per unit of product j procured from the supplier. If orders to the distributor for each product are independent, identically distributed random variables[7], the distributor then minimizes its inventory holding plus shortage costs for each product j by using a simple (stationary) base-stock policy with a target base-stock level S_j (see Nahmias 2000). Letting x_j denote the random variable for weekly orders received by the distributor for product j (with pdf $f(x_j)$), the single-period inventory cost equation at the distributor for product j, $L_d(S_j)$, given a starting inventory of I_j, is given by

$$L_d(S_j) = c_{dj}(S_j - I_j) + h_{dj}\int_0^{S_j}(S_j - x_j)f(x_j)dx_j + e_{dj}\int_{S_j}^{\infty}(x_j - S_j)f(x_j)dx_j. \quad (9.14)$$

[7]Although the assumption that successive orders to the distributor are statistically independent holds for our "base model" under no coordination, it does not necessarily hold under the coordination mechanisms we propose in the following subsections.

Equation (9.14) is a standard newsvendor equation, which is convex in S_j (see Nahmias, 2000). We next consider the relevant costs incurred by the retail stores.

Store inventory and shortage costs

Based on patterns observed from past POS data for the manufacturer's top three product lines, we assume that weekly (end-customer) demand at the stores for SKUs are stationary random variables. Each store follows a periodic review policy, with a period length of one week. Store inventory holding costs are assessed against end-of-period inventory in each period. We assume that if a store runs out of inventory for a particular SKU before receiving its next delivery, then it can backlog demand (at a per unit backlogging cost). The expression for each store's expected inventory holding plus backlogging cost is the same as Equation (9.14), except that the store's holding, purchasing, and shortage cost and demand parameters replace those of the distributor. Under our cost assumptions, if stores wish to minimize their inventory and backorder costs, they should follow a simple order-up-to or base stock policy[8]; we refer to this policy as the *base model.*

We next discuss two heuristic approaches for minimizing total expected system cost through mechanisms that reduce order variability at the distributor-retailer interface. Ideally, we would like to determine store-ordering policies that minimize total system (Distributor + Stores) cost. But, finding the optimal store ordering policy under our model is a highly complex problem. Therefore, we adopt two alternative (heuristic) approaches. In the first *penalty-based* approach, we assign a (possibly fictitious) penalty when actual store orders deviate from the mean order quantity. We then determine the policy that minimizes the stores' inventory and backorder costs (not system-wide costs) plus penalty costs, and compute the total system cost for this policy. The second approach, called the *policy-driven* approach, consists of selecting policies that are known to control or dampen upstream variance–such as a policy of ordering an amount equal to the moving average of past weekly demands. For each family of such order damping policies, we determine the optimal parameter settings that minimize total system costs.

4.2.2 Penalty-based approach. Consider the weekly orders that a store places for single item (or SKU). Let q be the random variable denoting a store's order quantity each week (for a single SKU), and

[8]Transactional data analysis indicates that stores do not always follow this policy, suggesting further cost reduction opportunities.

let $\mu = E(q)$ be the mean order quantity per week. Consider a *penalty* function $f(q-\mu)$ to discourage deviations of actual order quantities from the mean. For instance, for a given penalty parameter p, we might consider $f(q-\mu)$ to be either the absolute deviation function $p|q_n - \mu|$ or the squared deviation function $p(q_n - \mu)^2$. Both functions penalize positive as well as negative deviations of actual order quantities from their mean. We are interested in determining the optimal inventory (ordering) policy that a store must follow in order to minimize expected weekly inventory, backlogging, and penalty costs, assuming that end-customer demand follows a given stationary demand distribution. To motivate our penalty-based approach, consider the following general formulation of a (upstream) variance-constrained distribution problem (VCDP). Let ν denote a preferred maximum level of store order variance, and consider the problem:

[VCDP]

$$\textit{minimize } E[\{\textit{Distributor Inv.} + \textit{Transp. Cost}\} + \{\textit{Store Inv. Cost}\}]$$

subject to:

$$\textit{Order variability limit}: \qquad E[(q-\mu)^2] \leq \nu. \tag{9.15}$$

As ν decreases, we would expect distributor cost per period to decrease and store cost per period to increase. Although this model is not analytically tractable, formulation [VCDP] provides insight into methods for solving this problem. Stores' policy decisions translate into the distributor's demand process, resulting in a complex nonlinear program, particularly when store-order variance is a decision variable (see Geunes, 1999). Dualizing Constraint (9.15) in formulation [VCDP] (by multiplying it by a Lagrangian multiplier p and adding it to the objective function) results in the term $p(q_n - \mu)^2$ within the expected value expression in the objective function. As a heuristic approach, we next focus on minimizing those costs in the objective function (after dualizing the constraint) incurred by the store, i.e., the expected inventory holding, backlogging, and penalty costs. The penalty function serves as a surrogate for the cost burden that store-order variability imposes on the distributor.

Let I_n denote the store's on-hand inventory in period n (we number the periods in reverse) prior to receiving a replenishment, and let y_n denote the store's beginning inventory level in period n following replenishment. If $L(y_n)$ denotes the store's single-period expected purchasing, holding, and shortage costs, given a starting inventory of y_n, the minimum expected n-period *store cost* (including the added variability penalty function) as a function of the period n order quantity, q_n, becomes

$$g_n(I_n) = \min_{y_n \geq I_n} \{L(y_n) + f(q_n - \mu) + \beta E[g_{n-1}(y_n - x_n)\}, \tag{9.16}$$

where $q_n = y_n - I_n$. Equation (9.16) is a recursive equation giving the minimum expected n-period cost, $g_n(I_n)$, as a function of the initial inventory I_n. The third term in Equation (9.16) reflects the minimum discounted expected cost for period n - 1 onwards, where β is a discount factor.

Geunes (1999) characterizes the structure of the store's optimal policy for minimizing Equation (9.16) (over an infinite horizon) under the absolute value penalty, and provides a method to find the optimal parameters of this policy. This policy structure, called a*finite-generalizedbase-stock policy* (see Porteus, 1990, and Henig et al., 1997) involves setting two base-stock levels, S_1 and S_2 (where $S_2 \geq S_1$) and requires ordering either: up to S_1, up to S_2, or precisely μ depending on the inventory position before ordering. When compared to a standard base-stock policy (our base model), the finite-generalized base-stock policy leads to lower upstream order variability, since the store orders the fixed quantity μ in certain periods (assuming $S_1 \neq S_2$; otherwise this policy reduces to a standard order-up-to policy). The decrease in order variability seen by the distributor is clearly a function of the value of the deviation premium, p. When $p = 0$, the store follows its optimal base-stock policy and the distributor observes the same variability as the store. When $p = \infty$, the store absorbs all variability costs by always ordering μ, and the distributor therefore sees no variability. Tuning the value of p allows us to find better levels of order variability from a system-wide view, i.e., levels of store-order variability that lead to lower system costs compared to system costs without any coordination mechanisms. Note that this approach generalizes the approach proposed by Henig et al. (1997), which applies a premium only to quantities exceeding a contracted value between supplier and customer.

4.2.3 Policy-based approach. Under the policy-based approach, we assume that stores agree to implement specific variance-reducing inventory replenishment rules. The most promising among such policy-based mechanisms is the *weighted moving average (WMA) mechanism,* due to its ease of use and tractable mathematical analysis. Under a two-period weighted moving average policy, for example, the store order quantity consists of a weighted average of the last two demand realizations (Geunes, 1999, extends this scheme to an m-period WMA).

We use a single-product, single-store case to illustrate this mechanism, although the model readily extends to multiple stores and products (assuming product and store demands are all independent) as in the building-products supply chain. Let d_t denote the end-customer demand seen by the retail store in period t. Given a smoothing parameter α between 0 and 1, the two-period WMA policy determines the store's order quantity for period t as:

$$q_t = \alpha d_{t-1} + (1-\alpha) d_{t-2}. \tag{9.17}$$

Note that, by setting $\alpha = 1$, we obtain the base stock policy as a special case of the WMA policy. The variance of the store's order quantity, σ_q^2, for this policy is:

$$\sigma_q^2 = (\alpha^2 + (1-\alpha)^2)\sigma^2. \tag{9.18}$$

Equation (9.18) shows that a two-period WMA policy decreases the variance of order quantities facing the distributor compared to the variance of σ^2 under the base model. Changing the value of a allows us to dampen the level of order variability. Observe that since $q_t = \alpha d_{t-1} + (1-\alpha)d_{t-2}$ and $q_{t-1} = \alpha d_{t-2} + (1-\alpha)d_{t-3}$, the store's successive orders to the distributor are autocorrelated under the WMA policy mechanism. If, however, the distributor knows the value of d_{t-2} in period $t-1$ (when it places a stock replenishment order that will arrive at the beginning of period t), the $(1-\alpha)d_{t-2}$ part of the future store order, q_t, is known to the distributor at the time it places an order[9]. The only stochastic component of the future order q_t is the αd_{t-1} portion of the order. From the distributor's point of view, each store order contains both a deterministic component and a stochastic component. The stochastic order components are independent, stationary random variables with mean $\alpha\mu$ and variance $\alpha^2\sigma^2$, leading to an even greater store order variance reduction than indicated by Equation (9.18).

We next summarize selected computational results using our system model and actual supply chain transactional data[10]. Using the base model as the benchmark for comparison, we evaluate the cost performance of the proposed variance-damping mechanisms. In particular, we discuss the impacts of using the two-period WMA policy assuming that

[9]The distributor can obtain the values of past store demands through either information sharing or by reconstructing successive demands based on knowledge of the store's ordering policy combined with a record of past store orders.

[10]Note that the results presented here assume that the distributor uses a stationary base-stock level for each product, which is not necessarily optimal under an autocorrelated order process, which occurs under our proposed coordination mechanisms. Because of this we can view the results presented here as providing a lower bound on potential savings.

all stores apply the same damping parameter, α (for test results on the penalty-based approach under randomly generated problem instances, see Geunes, 1999)[11]. We partitioned the set of 62 stores into 18 customer clusters. Although the distributor stocks and supplies products belonging to several of the manufacturer's product lines, we focus on the three major product lines (containing 248 SKUs) that account for more than 85% of distribution by weight. We obtained transportation and inventory cost parameters based on estimates provided by the manufacturer. By evaluating total system cost at various values of α between 0 and 1, we found that the minimum total expected system cost occurred at $\alpha = 0.8$, which provides a 68% reduction in variance at the distributor-store interface (using Equation 9.18). Compared to the base model, the simple two-period WMA policy provided approximately 2.76% savings in total expected system cost; extrapolating these savings to all distributors in the manufacturer's nationwide network implies system savings in excess of $1M annually. Past literature on the bullwhip effect emphasizes the need to lessen the amount of demand amplification at upstream stages in the supply chain. Our results support this conclusion and show that supply chain costs can be further reduced if variability is dampened at the trailing end of the supply chain (i.e., at the retail store level).

5. Concluding Remarks

This paper has discussed three interesting and new models, applicable to practical distribution problem settings, that arose through a joint project with a leading US building-products manufacturer. Our analysis of the distribution system came at an opportune time, following a major change in the architecture and performance requirements of the system. These major system transformations resulted from a significant change in the purchasing patterns of US consumers over the past two decades. Over this period, consumers shifted from traditional independent retailers to large retail chains in many industries because of the combination of price, product quality, and service these retail chains provide. Specifically, the problems we consider are motivated by the new challenges that the big-box retailers presented to the large building-products manufacturer. Coping with these novel distribution challenges requires researchers and industry practitioners to develop new models and solution techniques.

[11]System costs can decrease further by using product- and store-specific damping parameters; however, in this case, the solution space dramatically increases due to the increased number of decision variables.

The problems we address arise at different points in the supply chain: the *fee-setting* problem occurs at the manufacturer-distributor interface, the *delivery planning under uncertain demands* problem arises within the distributor's operations, and the *upstream variability control* problem relates to the distributor-retailer interface. Although we can isolate the particular supply chain stages where each problem arises, we emphasize that performance improvements at these isolated points will often provide indirect benefits throughout the chain. For example, reducing order variability at the distributor-retailer interface not only leads to lower costs at these stages, but also leads to less compensation paid out by the manufacturer, since distributor costs decrease. As distribution channels continue to evolve, particularly through the use of advanced information technologies, new and interesting distribution planning problems and modeling opportunities will likely continue to arise.

Acknowledgments

This work was partially supported by the Institute for the Study of Business Markets at Penn State University, University Park, PA.

References

Baganha, M.P., and M.A. Cohen, 1998. "The Stabilizing Effect of Inventory in Supply Chains," *Operations Research,* 46(Su 3), S72-S83.

Balakrishnan, A., H.P. Natarajan, and M.S. Pangburn, 2000 "Optimizing Delivery Fees for a Network of Distributors," *Manufacturing & Service Operations Management,* 2(3), 297-316.

Bertsimas, D.J., and D. Simchi-Levi, 1996. "A New Generation of Vehicle Routing Research: Robust Algorithms, Addressing Uncertainty," *Operations Research,* 44(2), 286-304.

Birge, J.R., and F. Louveaux, 1997. *Introduction to Stochastic Programming,* Springer, New York, NY.

Cachon, G.P., 1999. "Managing Supply Chain Demand Variability With Scheduled Ordering Policies," *Management Science,* 45(6), 843-856.

Cachon, G.P., and P.H. Zipkin, 1999. "Competitive And Cooperative Inventory Policies In A Two-Stage Supply Chain," *Management Science,* 45(7), 936-953.

Chen, F., 1999. "Decentralized Supply Chains Subject to Information Delays," *Management Science,* 45(8), 1076-1090.

Clark, A.J., and H. Scarf, 1960. "Optimal Policies for a Multi-Echelon Inventory Problem," *Management Science,* 6, 475-490.

Gavirneni, S., R. Kapuscinski, and S. Tayur, 1999. "The Value Of Information In Capacitated Supply Chains," *Management Science,* 45(1), 16-24.

Gendreau, M., G. Laporte, and R. Seguin, 1996. "A Tabu Search Heuristic For The Vehicle Routing Problem With Stochastic Demands And Customers," *Operations Research,* 44(3), 469-477.

Gendreau, M., Laporte, G., and Seguin, R., 1996. "Stochastic Vehicle Routing," *European Journal of Operational Research,* 88(1), 3-12.

Geunes, J., 1999. *Models To Optimize Multi-Stage Linkages In Manufacturing And Distribution Operations,* Ph.D. Thesis, Penn State University.

Goldberg, D.E., 1989. *Genetic Algorithms in Search, Optimization, and Machine Learning,* Addison-Wesley, Reading, MA.

Henig, M., Y. Gerchak, R. Ernst, and D. F. Pyke, 1997. "An Inventory Model Embedded in Designing a Supply Contract," *Management Science,* 43(2), 184-189.

Higle, J.L., and S. Sen, 1996. *Stochastic Decomposition: A Statistical Method For Large Scale Stochastic Linear Programming,* Kluwer, Boston, MA.

Lee, H.L., V. Padmanabhan, and S. Whang, 1997. "Information Distortion in a Supply Chain: The Bullwhip Effect," *Management Science,* 43(4), 546-558.

Lee, H.L., K.C. So, and C.S. Tang, 2000. "The Value of Information Sharing in a Two-Level Supply Chain," *Management Science,* 46(5), 626-643.

Lee, H.L., and S. Whang, 1999. "Decentralized Multi-Echelon Supply Chains: Incentives And Information," *Management Science,* 45(5), 633-640.

Li, C., and P. Kouvelis, 1999. "Flexible and Risk-Sharing Supply Contracts Under Price Uncertainty," *Management Science,* 45(10), 1378-1398.

Moinzadeh, K., and S. Nahmias, 2000. "Adjustment Strategies for a Fixed Delivery Contract," *Operations Research,* 48(3).

Nahmias, S., 2000. *Production and Operations Analysis, 4th ed.,* McGraw-Hill, Chicago, IL.

Porteus, E.L., 1990. "Stochastic Inventory Theory," Ch. 12 in D.P. Heyman and M. Sobel (eds.), *Handbooks in OR & MS, v2,* North-Holland: Elsevier Publishers B.V., New York, NY.

Porteus, E.L., 2000. "Responsibility Tokens in Supply Chain Management," *Manufacturing & Service Operations Management,* 2(2), 203-219.

Simchi-Levi, D., P. Kaminsky, and E. Simchi-Levi, 2000. *Designing and Managing the Supply Chain,* Irwin McGraw-Hill, Boston, MA.

Stewart, W.R., Golden, B.L., 1983. "Stochastic Vehicle Routing: A Comprehensive Approach," *European Journal of Operational Research,* 14(4), 371-385.

Tsay, A., 1999 "The Quantity Flexibility Contract and Supplier-Customer Incentives," *Management Science,* 45(10), 1339-1358.

III

MODELS AND APPLICATIONS FOR SUP-PLY CHAIN PLANNING AND DESIGN

Chapter 10

A MATHEMATICAL PROGRAMMING MODEL FOR GLOBAL SUPPLY CHAIN MANAGEMENT: CONCEPTUAL APPROACH AND MANAGERIAL INSIGHTS

Panos Kouvelis
John M. Olin School of Business
Washington University
St. Louis, MO 63130
kouvelis@olin.wustl.edu

Meir J. Rosenblatt
Faculty of Industrial Engineering and Management
Technion - Israel Institute of Technology
Haifa 32000
Israel
meirr@ie.technion.ac.il

Abstract In this paper, we study the design of global facility networks. We present a mixed integer programming model that captures essential design tradeoffs of such networks and explicitly incorporates government subsidies trade tariffs and taxation issues. The resulting formulation can be solved for reasonable size problems with commercially available mathematical programming software. Focusing on special cases of the problem enables us to provide useful insights on preferable international facility networks for various environments. We demonstrate the pervasive, and often dominating, effects of subsidized financing, tariffs, regional trade rules, and taxation in shaping the manufacturing and distribution network of global firms.

J. Geunes et al. (eds.),
Supply Chain Management: Models, Applications, and Research Directions, 245–277.

1. Introduction

1.1 Different Approaches in Structuring Global Networks

Structuring global manufacturing and distribution networks is a complicated decision making process. The typical input to such a process is a set of markets to serve, a set of products that the company will produce and sell, demand projections for the different markets, and information about future macroeconomic conditions, transportation and production costs. Given the above information, companies have to decide, among other things, where to locate factories, how to allocate production activities to the various facilities of the network, and how to manage the distribution of products (e.g., where to locate distribution facilities). There are three main approaches in structuring global facility networks: *(1) Product Family-Focused*: where plants may be located in different parts of the world and each specializes in a specific product family (a set of products with similar process and market characteristics). Each product family plant is essentially an independent small company that supplies its product group to all markets in which the product is sold. *(2) Process Focused*: where individual plants are typically dedicated to performing specific process steps for a variety of different products. Sometimes a product is produced entirely by a single plant, but more often the plant is only one of several in the process chain. *(3) Market-Focused*: where plants are located at markets that they plan to serve. All products sold in that market are served locally. Each market is essentially insulated from other markets, and plants serve the needs of the local market only.

Most global facility networks firms use are hybrids of the above approaches. For example, ***Product/Process Focused*** networks have plant subnetworks that produce specific product families. The products within a product family can be completed in more than one facility of this subnetwork, and their various subassemblies are allocated to the various plants of the subnetwork. In other words, the plants within the subnetwork are process focused, but the subnetworks are product focused.

1.2 The Importance of Facility Financing and Corporate Taxation

Among the most important factors distinguishing the organization of global production activities from strictly domestic ones are the influences of various governmental policies. Governments often use special financing and taxation incentives to stimulate production investments in their country. The most important of such subsidies are:

- **Cash grants**: usually given as an incentive for investment either in a particular development region or for development of a particular product or industry. Northern Ireland offers tax free grants of up to 50 percent of the cost of new factory buildings, machinery and equipment. Of a similar nature is Israel's Capital Investment Encouragement Law where firms can receive cash subsidies up to 38 percent of the project's cost. Israel has recently provided Intel with a 600 million dollar cash grant for a 1.6 billion dollar new facility.

- **Loans at reduced interest rates**: the size of these loans can be related to the total investment capital and may vary both in size and interest rate according to the importance of the region or the project. For example, in Austria the European Recovery Programme provides loans at interest rates significantly below market levels for several types of investments, and loans for large projects could amount to up to 50% of the total investments.

- **Taxation Related Incentives**: for locating in high priority regions or manufacturing products that the government is trying to foster, firms might be exempt from paying income tax or pay lower rates for a number of years. For example, in Brazil reduced income tax rates are available in the Monaus region – Amazon forest, while in Northern Ireland the corporate income tax rate is among the lowest in the world.

In this paper we present a mathematical modeling framework that explicitly considers financing and corporate taxation issues in the design of ***Hybrid Product/Process Focused*** facility networks. The multi-period model maximizes the discounted after-tax cash flows of the firm. The paper maintains a modeling rather than an algorithmic focus.

1.3 Literature Review

One of the early papers in this area, which presents a single period model for determining simultaneously international plant location and financing decisions under uncertainty, is by Hodder and Dincer [1986]. The model uses a mean-variance approach in the objective function to incorporate price/exchange rate uncertainty and risk aversion in the location problem, an approach that builds on earlier work of Jucker and Carlson [1976] and Hodder and Jucker [1985a, b]. Aspects of subsidized government financing, tariffs and taxation are not modeled.

A succession of papers by Cohen and Lee address a variety of issues of facility network design in a global context. Cohen and Lee [1988] deal with intra-firm international trade issues. In their model, plants are specialized units that produce the sub-assemblies of the final product and the Distribution Centers (DCs) perform the final assembly. Plants are charged with fixed and variable costs, while DCs are charged only with variable costs. Cohen and Lee [1989] developed a normative integer programming model for resource deployment decisions in a global manufacturing and distribution network for a U.S. based manufacturer of personal computers. Their model enables estimation of before and after-tax profitability, including exchange rate effects to a numeraire currency. Their implementation is in GAMS/MINOS (Brooke, Kendrick and Meeraus 1988), which has no integer programming capability. Consequently, they only solve the continuous versions of their models, pre-specifying "alternate sets of integer decisions variables." Cohen, Fisher and Jaikumar [1989] presented a multiperiod extension of the above model, which explores tradeoffs between centralization and localization of supply chain decision making. A hierarchical procedure is proposed as a heuristic solution approach to the problem. Huchzermeier and Cohen [1996] present a modeling framework that integrates the network flow and option valuation approaches to global supply chain modeling. The model maximizes discounted, expected, global after-tax profit for a multinational firm in terms of numeraire currency. The results in the paper provide a methodology for quantifying the risks and returns of flexible global manufacturing strategies. Their work clearly demonstrates how flexible facility networks with excess capacity can provide real options to hedge exchange rate fluctuations in the long term. This work does not emphasize explicit modeling of facility financing and tax incentives and the resulting implications for the global network structure. Finally, various company-specific facility network design issues and decision support systems have been discussed in the applications-oriented literature, see Breitman and Lucas [1987] work on General Motors, Arntzen et al. [1995] work on Digital Equipment Corporation and Davis [1993] work on Hewlett-Packard.

Our work in this paper is along the research directions pointed out in the recent paper of Revelle and Laporte [1996]. As they point out, even though plant location and configuration of production/distribution networks have been studied for many years, a number of important real world issues have not received adequate attention. Even though there is ample anecdotal evidence on the strong effects of financing and taxation factors on the structure of global manufacturing and distribution networks (see Bartmess and Cerny [1993] and MacCormack et al. [1994]), there is a lack of reasonably simple, yet comprehensive, models which illustrate the effects of such factors on global location/network configuration decisions. Our research in this paper attempts to fill this gap.

The main contributions of this paper are:

a) the incorporation of such issues as government subsidies in facility financing, trade tariffs and taxation in a manageable size linear integer program that captures the multiperiod nature of international plant location and distribution network configuration decisions;

b) a model which explicitly captures intra-firm material flows (from plants to DCs) and provides insights on preferable international facility network structure for various environments;

c) the illustration of the dominating effects of subsidized financing, tariff and local content rules, and taxation in shaping the optimal structure of global facility networks.

The structure of this paper is as follows: In Section 2 we present a mixed integer programming formulation of the global facility network design problem explicitly accounting for government financing and taxation issues, and report on the computational experience with the model via standard mathematical programming software. Section 3 uses a special case of the model ("identical countries") to provide insights on the effects of financing, taxation, regional trading zones and local content rules on the structure of global facility networks. Use of the mixed integer programming model in an illustrative international location decision in Section 4 further demonstrates the pervasive effects of the above factors in global facility network design. Many of the managerial insights obtained in the "identical countries" special case reappear in the realistic environment of our illustrative decision instance.

2. Problem Statement and Formulation

2.1 Problem Statement

A firm uses a Hybrid Product/Process Focus approach in designing its global network of plants and distribution centers. The firm plans to produce a new product with sales expectations in many countries. The product is composed of many different subassemblies, but requires only one unit per subassembly (i.e., a two level product tree, with level 0 the final product and level 1 all its subassemblies). A network of facilities will be developed dedicated to the production of this product, and each one of the plants within the network will be dedicated to the production of a specific subassembly. In order to assemble the final product, distribution centers (DCs) will be located in various countries. To influence the firm's location decision, various governments are willing to grant to it loans with subsidized interest rates and tax incentives that reduce the facility costs. The firm wants to develop a network of plants and DCs that maximizes its discounted after-tax profit over the planning horizon for this product.

2.2 Model Assumptions

- There are two facility levels: plants (producing subassemblies) and distribution centers (assembling the final product).
- Plant and DC locations are selected among an identified set of candidate locations.
- The plants and DCs remain open throughout the finite planning horizon.
- The firm is a price taker in each market. All prices are quoted in the currency of the market the product is sold and then translated into a common currency (say $) by using the real exchange rate.
- The market demand and the selling price are independent of the structure of the facility network.
- At most one Distribution Center (DC) is allowed in each country. The main rationale behind this assumption is that since shipment cost of a subassembly or a final product within a country is assumed constant in our model, the opening of a second DC will result in additional fixed costs with no transportation cost savings.
- Demand and supply clear in each time period and hence there are no inventory costs.
- Corporate income tax is paid on profit in each country of operation.

2.3 Notation

Indices

i: subassembly index, $i = 1, \ldots, I$;

n: country of operation index, $n = 1, \ldots, N$;

k: distribution center (DC) (as well as country that DC is located) index, $k = 1, \ldots, N$;

m: country (potential plant location) index, $m = 1, \ldots, N$;

j: country (market) index, $j = 1, \ldots, J$;

t: time period, $t = 1, \ldots, T$;

Parameters

f_{imt}: fixed cost of plant that produces subassembly i in country m in period t;

v_{imt}: unit variable production cost of subassembly i produced in country m in period t;

F_{kt}: fixed cost of a DC in country k in period t;

V_{kt}: unit variable assembly cost of a DC in country k in period t;

S_{kjt}: cost to ship one unit of the final product from a DC in country k to market j in period t (shipment cost calculations include any assessed trade tariffs and other duties);

$\vartheta_i S_{mkt}$: cost to ship one unit of subassembly i from country m to country k in period t (i.e., subassembly transportation cost is expressed as a fraction of the transportation cost of the final product, with ϑ_i the fraction for subassembly i);

D_{jt}: demand of the final product in country j in period t;

r_n: per-period interest rate on the loan in country n;

$R_t(r, T)$: capital recovery factor for period t given interest rate r and planning horizon T;

$\rho_t(r, T)$: interest calculation factor for period t given interest rate r and planning horizon T;

A_{im}: required investment to build a plant for subassembly i in country m;

W_k: required investment to build a DC in country k;

t_{nt}: marginal corporate income tax rate in country n in period t;

P_{jt}: sales price of the final product (translated into \$ with a real exchange rate) in country j in period t;

p_{imkt}: transfer price of subassembly i made in country m shipped to DC in k in period t;

d_{nt}: applicable depreciation rate in country n in period t (% per period);

K_{im}: capacity of a plant producing subassembly i in country m;

C_k: capacity of a DC in country k;

β_{nt}: discount rate of after-tax cash flows in country n in period t;

B_n: maximum loan that country n can give to the firm;

Decision Variables

X_{imkt}: units of subassembly i produced in country m and assembled in country k in period t;

Y_{njt}: units assembled in country n and used to satisfy demand in market j in period t;

w_k: the loan that country k government will grant to the company to build a DC;

ψ_m: the loan that country m government will grant to the company to build subassembly plants;

y_{im}: $= 1$ if a plant to produce subassembly i is located in country m, and 0 otherwise;

z_k: $= 1$ if a DC is operated in country k, and 0 otherwise.

2.4 Basic Components of the Model

2.4.1 The Objective Function. Revenue of units assembled (α_{nt}) and subassemblies produced (μ_{nt}) in country n in period t, where:

$$\alpha_{nt} = \sum_{j=1}^{J} P_{jt} Y_{njt} \text{ and } \mu_{nt} = \sum_{i=1}^{I} \sum_{k=1}^{N} p_{inkt} X_{inkt}.$$

Transportation costs from a DC (a_{nt}) and subassembly plants (b_{nt}) located in country n in period t, where:

$$a_{nt} = \sum_{j=1}^{J} S_{njt} Y_{njt} \text{ and } b_{nt} = \sum_{i=1}^{I} \sum_{k=1}^{N} \vartheta_i S_{nkt} X_{inkt}.$$

Annual fixed costs of a DC (δ_{nt}) and subassembly plants (γ_{nt}) in country n in period t, where:

$$\delta_{nt} = z_n F_{nt} \text{ and } \gamma_{nt} = \sum_{i=1}^{I} y_{in} f_{int}.$$

Cost of goods sold for the final product assembled at the DC (c_{nt}) and assembly related costs of a DC (φ_{nt}) in country n in period t, where:

$$c_{nt} = \sum_{i=1}^{I} \sum_{m=1}^{N} p_{imnt} X_{imnt} \text{ and } \varphi_{nt} = V_{nt} \sum_{j=1}^{J} Y_{njt}.$$

Variable production costs of subassembly plants in country n in period t,

$$e_{nt} = \sum_{i=1}^{I} v_{\text{int}} \left(\sum_{k=1}^{N} X_{inkt} \right).$$

Loan payment in period t (g_{nt}) and interest payment (g'_{nt}) for loan given by country n for the construction of subassembly plants in its territory, where:

$$g_{nt} = \psi_n R_t (r_n, T) \text{ and } g'_{nt} = \psi_n \rho_t (r_n, T).$$

Loan payment in period t (h_{nt}) and interest payment (h'_{nt}) for loan given by country n for the construction of a DC in its territory, where:

$$h_{nt} = w_n R_t (r_n, T) \text{ and } h'_{nt} = w_n \rho_t (r_n, T).$$

Depreciation expense in country n in period t,

$$\tau_{nt} = \left(\sum_{i=1}^{I} A_{in} y_{in} + W_n z_n \right) d_{nt}.$$

Before-tax income in country n in period t,

$$\pi_{nt} = [\alpha_{nt} + \mu_{nt} - (a_{nt} + b_{nt} + \gamma_{nt} + \delta_{nt} + c_{nt} + e_{nt} + \varphi_{nt} + g'_{nt} + h'_{nt} + \tau_{nt})].$$

Corporate income tax paid in country n in period t,

$$CT_{nt} = \pi_{nt} t_{nt}.$$

Cash expenditures in fixed assets in year 0 that are not financed by external sources.

$$IN = \sum_{m=1}^{N} \left[\left(\sum_{i=1}^{I} A_{im} y_{im} \right) - \psi_m \right] + \sum_{k=1}^{N} (W_k z_k - w_k).$$

So the objective function which maximizes the net present value is:

$$\text{OBF} = \max \left[\frac{\sum_{n=1}^{N} \sum_{t=1}^{T} (\pi_{nt} + g'_{nt} + h'_{nt} + \tau_{nt} - g_{nt} - h_{nt} - CT_{nt})}{(1 + \beta_{nt})^t - IN} \right].$$

2.4.2 The Set of Constraints. The Constraints are:

1) Demand Constraints: $\sum_{n=1}^{N} Y_{njt} \leq D_{jt},\ \forall\ j, t$

2) Plant Capacity Constraints: $\sum_{k=1}^{N} X_{imkt} \leq K_{im},\ \forall\ i, m, t$

3) DC Capacity Constraints: $\sum_{j=1}^{J} Y_{kjt} \leq C_k,\ \forall k, t$

4) Conservation of Subassembly Flows: $\sum_{m=1}^{N} X_{imkt} = \sum_{j=1}^{J} Y_{kjt},\ \forall\ i, k, t$

5) Country Budget Constraints: $\psi_n + w_n \leq B_n,\ \forall n$

6) Compatibility of Decision Variable Values constraints:
$\sum_{k=1}^{N} \sum_{t=1}^{T} X_{imkt} \leq y_{im} \cdot M,\ \forall\ i, m$ and $\sum_{j=1}^{J} \sum_{t=1}^{T} Y_{kjt} \leq z_k \cdot M,\ \forall\ k,$
where M is a large number such that $M \geq \sum_{j=1}^{J} \sum_{t=1}^{T} D_{jt}$.

7) Nonnegative Profit in Each Country in Each Period (assumption of convenience to avoid unnecessarily complicated tax calculations): $\pi_{nt} \geq 0,\ \forall\ n, t.$

8) Loan Ceilings: $\psi_m \leq \sum_{i=1}^{I} A_{im} y_{im},\ \forall\ m$ and $w_k \leq W_k z_k,\ \forall\ k.$

2.5 Computational Experience with Proposed Model

The following table reports computational results for solving various size problems. Table 10.1 provides sizes of our model with the use of standard mathematical programming software (specifically GAMS/OSL2) on a Cyrix 686-P90 processor.

We would not, however, want to give the false impression that substantially larger size problems can be solved on standard PCs. Larger size formulations (in our experience up to $I = 10$, $N = 10$, $J = 100$, $T = 10$) can be solved on minicomputers, (e.g., VAX 6400) with the use of standard mathematical programming software in less than an hour of CPU time.

Number of Subassemblies (I)	Number of Candidate Location Countries (N)	Number of Markets (J)	Number of Time Periods (T)	Solution Times (seconds)
3	6	50	5	14
5	6	100	5	41
20	6	100	5	106
5	20	100	5	140
5	25	50	5	108
3	35	35	5	121
5	10	250	5	187
5	6	100	15	257
5	10	50	15	273
10	15	50	5	87
10	10	100	5	104

Table 10.1. Computation Time for Various Size Markets

3. Special Case: Identical Countries

3.1 Base Case: Tariff and Transportation Cost Considerations

In this section we analyze a special case of our model to obtain useful insights into the nature of location decisions in a global environment. We have relaxed the capacity constraints on plants and DCs (i.e., uncapacitated situation) and assume straight-line depreciation. Initially, we assume no financing and equal tax rates across countries (we refer to this as the "base case").

We consider "identical countries" and "identical periods" (i.e., $D_{jt} = D, P_{jt} = P, t_{jt} = t,$ and $\beta_{jt} = \beta$ for all j) with a trade tariff assessed on flows crossing their borders. The product consists of I "identical subassemblies" in terms of revenue and cost parameters in our model. For example, $\vartheta_i = \vartheta$ for all i. The assumption is not intended to imply physical similarity but only similarity of cost parameters. We assume that only one plant of each subassembly type will be built (due to, for example, economies of scale). Also, there is no differentiation among subassembly plants or DCs (we use the term "identical plants or DCs" with an analogous meaning) and specifically

$$A_{im} = A, v_{im} = v, f_{im} = f \text{ for } i = 1, 2, \ldots, I \text{ and } m = 1, 2, \ldots, N;$$

$$\textit{and } W_k = W, V_k = V, F_k = F \text{ for } k = 1, 2, \ldots N.$$

Without loss of generality, we also assume that a DC can be constructed in every market (i.e., $N = J$). The transportation costs within the countries are C per unit of final product (i.e., $S_{kk} = C$); however, they increase by the assessed trade tariff whenever borders are crossed. Specifically, transportation costs per unit of final product across borders equal $S_{kj} = C(1+\Delta)$, $\Delta > 0$, for all $(k, j), k \neq j$. We assume that the transfer prices p are the same for all subassemblies and based on a world-wide market standard that is closely monitored by governments (i.e., the firm cannot change transfer prices to take advantage of tax conditions). Finally, we assume that prices and costs are structured such that the firm will always show nonnegative profit in every country as long as a DC in any country serves at least the total demand for the country of its location.

Proposition 10.1 shows that we can ignore all cases where subassembly plants are located in different countries.

Proposition 10.1 *Given I identical plants which can be constructed in any of N identical countries, it is preferable, or at least as profitable as any other configuration, to build all I plants in one country.*

Thus, it makes sense for all subassembly plants to be located in one country, but we still must determine how many DCs to build. We examine two cases where the number of DCs equal to 1 or q (where $1 < q \leq N$). The Net Present Values (NPVs) of these configurations are given below, and the DC locations that maximize the NPV are assumed (i.e., markets that do not contain a DC are all served from the same DC located in the country containing the subassemblies):

$$NPV(1\ DC) = -IA - W + \sum_{\omega=1}^{T} \{(1-t)[NPD - (N + N\Delta - \Delta)CD$$
$$-NI\vartheta CD - If - F - NIvD - NDV] + \tfrac{t(IA+W}{T}\}/(1+\beta)^{\omega},$$

After rearranging terms, the NPV (q DCs) is equal to:

$$\text{NPV}\,(q\ \text{DCs}) = -IA - qW + \sum_{\omega=1}^{T} \{(1-t)\,[NPD$$
$$+CD(I\vartheta\Delta(1-q) + \Delta(q-N) - N(I\vartheta+1)) - If - qF - NIvD - NDV]$$
$$+\tfrac{t(IA+qW)}{T}\}/(1+\beta)^{\omega}.$$

The following proposition provides insights on the nature of the DC configuration for this special case.

Proposition 10.2 *For the case of identical plants and identical countries,*

(a) It is preferable to either build a single DC in the country where all plants are located or to construct DCs in all countries;

(b) For the infinite time horizon case ($T \to \infty$), a single DC is optimal if either (1) $I\vartheta \geq 1$, or (2) $I\vartheta < 1$ and $D \leq D*$, where

$D* = [\beta W/(1-t) + F]/[\Delta C\,(1 - I\vartheta)]$; *otherwise a configuration with DCs in all countries is optimal.*

According to the above proposition, if it is more expensive to transport the subassemblies individually than to transport the final product $(I\vartheta > 1)$, then it is preferable to use a centralized distribution structure with the product transported from the country where it is manufactured and assembled to all of its final markets. If the condition is reversed, i.e., transportation of the final product is more expensive, then given a substantial demand in foreign markets, it is preferable to open a DC in each one of the markets (i.e., a decentralized distribution structure) and transport subassemblies from the subassembly manufacturing plants.

The analysis of this special case provides the following insights for environments where the transportation cost elements (in our case transportation cost differences were generated through the Δ factor, i.e., trade tariffs and other restrictive cross-border flow regulations) tend to dominate:

(a) when transportation of subassemblies is expensive, firms prefer a centralized manufacturing and distribution structure;

(b) as the transportation cost element between countries and/or cross-border flow government imposed penalties increase firms tend to open DCs in markets away from the manufacturing facilities; and

(c) governments can attract DC investments in such environments by offering financing help through loans.

3.2 Government Incentives: Financing and Taxation

Governments can attract facility investments in their country via subsidized financing of facility (plant and/or DC) construction, as the following proposition illustrates:

Proposition 10.3 *For the case of identical plants and identical countries, if only one country offers attractive financing for facility construction with interest rate* $r < \beta/(1-t)$, *then all I plants and a distribution center will be built in that country.*

As we demonstrate in Section 4, in many cases even relatively small subsidization of loan interest rates by the government is adequate to attract facility investments in the country. However, the level of facility investment in the country by a firm is non-decreasing with the level of subsidized financing provided by the government.

Next we explore the situation where one country charges a lower tax rate than the rate of other countries.

Lemma 10.4 *If there is only one DC, the plants and the DC are located in the country with the lower tax rate.*

Lemma 10.5 *If more than 1 DC is open, the plants and the DC serving any markets without a DC will be located in the country with the lower tax rate.*

Proposition 10.6 follows from the two lemmas.

Proposition 10.6 *For the case of identical plants and identical countries, if one country offers a lower tax rate (t') than the rest (t), then all I plants will be built in that country. In addition, it is preferable to either build a single DC in the lower tax rate country or to construct DCs in all countries.*

Let us observe the relevant expressions for NPV(1 DC) - NPV (q DCs) in the base case (see proof of Proposition 10.2) and lower tax rate country case (see proof of Proposition 10.6). The difference in these two expressions is the term $(t - t')\, D\left[P - Ip - V - C\left(1 + \Delta\right) + \Delta CI\vartheta\right]$, which is positive due to the assumption of the price structure ensuring nonnegative profits in each country of operation. This implies that it is easier to satisfy the inequality NPV(1 DC) - NPV (q DCs) > 0 in the lower tax rate country case than under the base case. Therefore, the firm is more likely to desire complete centralization of its distribution network when one country offers a lower tax rate than for the case where all countries offer the same tax rate.

According to Proposition 10.6 and the above observations, by lowering its tax rate, not only may a country entice firms to open plants and/or distribution centers, but also the amount of demand flowing through existing distribution centers in that country may increase. Country environments with differential tax rates tend to favor more centralized manufacturing and distribution network structures. Even rather small tax rate differences can induce firms to move their facility investments to countries with lower tax rates, as will become more apparent through our discussion in the next section.

3.3 Tariff Structure: Regional Trading Zones and Local Content Rules

Governments sometimes enter into multilateral trade agreements which allow for tariff-free trade among nations that belong to the pact. In this section we introduce trading pacts into the special case and describe the changes that the agreements may cause for a firm's global distribution network configuration. We begin by stating two lemmas.

Lemma 10.7 *For the case of identical plants and identical countries, if the N countries are divided into Z tariff-free trade zones, then no more than one DC should be located in each trading zone.*

Lemma 10.8 *For the case of identical plants and identical countries, if the N countries are divided into Z tariff-free trade zones, then all I plants should be built in any of the countries that form the trading zone which contains the most countries.*

Lemmas 10.7 and 10.8 allow us to construct the following proposition, which describes the impact that trading zones have on the degree of centralization of a firm's distribution network.

Proposition 10.9 *Consider the case of identical plants and identical countries, where the N countries are divided into Z tariff-free trade zones. If* $I\vartheta > 1\,(I\vartheta < 1)$*, then a completely centralized distribution network, with only one DC, is more likely (less likely) to be optimal than for the case without tariff-free zones.*

Proposition 10.9 implies that the centralization/decentralization distribution network configuration decision is more sensitive to the relative transportation cost of subassemblies vs. the final product when trading zones are introduced. Specifically, if it is more expensive to transport the subassemblies individually than to transport the final product $(I\vartheta > 1)$, then a centralized distribution is more likely than in the base case, as the penalty for transporting the final product from one location is smaller as the home (tariff-free) market is bigger. On the other hand, if transportation of the final product is more expensive $(I\vartheta < 1)$, then centralized distribution is less likely than in the base case, as the penalty for transporting subassemblies across borders is smaller as fewer are charged the tariff (i.e., more subassemblies are transported tariff free to the DC in the same zone).

The impact of trading zones can be summarized by stating that when trading zones are established and the cost of transporting subassemblies is not high compared to the final product, then the distribution network is likely to become "regionalized" instead of centralized in one country or decentralized among all markets. In other words, the network becomes decentralized over trading zones (instead of countries) but centralized within each zone. On the other hand, if economies of scale exist such that only one plant will be built for each subassembly, and if the total cost of transporting subassemblies is high compared to the final product, then the establishment of trading zones will more likely result in manufacturing and distribution centralization, i.e., all facilities in one location. In that case, the zone offering the best incentives may attract all of the firm's worldwide investment.

Many governments establish "local content rules," which require firms to have a certain portion of their final products originate from that country. The percentage requirement has been known to be as high as 90%, and the specific rules and penalties may vary greatly among countries (Munson and Rosenblatt 1997). Consider a local content rule which imposes a tariff Δ for domestic shipment from the distribution center unless the value of the subassemblies from domestic plants plus the value added at the distribution center is at least as high as the local content percentage times the final good's value. In other words, the country treats the final product as an imported good unless a certain portion of that good is produced in the country. This type of rule prohibits firms from using cheap foreign labor to produce the bulk of the product while adding little value in the country of sale.

Since the countries are identical under the base case scenario, if one government introduces a local content rule, then a DC and all I plants will be diverted to that country (because there would be no penalty for doing so). Therefore, we will examine the situation from Section 3.2 where one of the countries (A) has a lower tax rate than the other countries, and country B introduces a local content rule in order to attract investments. Now if conditions are such that, prior to the introduction of any local content rules, a completely centralized distribution network is optimal, then from Lemma 10.4, country A will have all of the facilities (including the only DC). The local content rule introduced by country B will have no effect because the firm currently has no DC in country B. Therefore, we will assume that conditions are such that each country has a DC, and (from Lemma 10.5), all I plants currently reside in country A. Country B introduces the local content rule to try to "force" some of the plants to move to that country.

Clearly, based on Lemma 10.5, the firm will want to move as few plants as possible from country A to country B in order to satisfy the local content requirement. The local content rule is defined here as $\Phi pD + VD \geq \Omega(IpD + VD)$, where Φ is the number of subassembly plants located in country B and Ω is the local content percentage $(0 \leq \Omega \leq 1)$. Therefore, the number of subassemblies that the firm has to move in order to satisfy local content is the smallest integer Φ^* such that

$$\Phi^* \geq [\Omega Ip - (1 - \Omega) V] / p.$$

Proposition 10.10 describes the conditions under which the firm should satisfy the local content requirement imposed by country B.

Proposition 10.10 *For the case of identical plants and identical countries with country A having a lower tax rate than the other countries, it is worthwhile for the firm to move Φ^* subassembly plants from country A to country B in order to satisfy country B's local content rule if:*

$$\Phi^* < \tag{10.1}$$
$$\frac{-W + \sum_{\omega=1}^{T} \ (1-t')\Delta CD + \frac{tW}{T} - (1-t)F - (t-t')D(P - Ip - C - V) \ /(1+\beta)^{\omega}}{\sum_{\omega=1}^{T}(t-t') \ NDp - (N-1)(1+\Delta)\vartheta CD - \vartheta CD - f - NDv - \frac{A}{T} \ /(1+\beta)^{\omega}}.$$

If the right-hand-side of (10.1) is greater than I, then country B can impose any size Ω (up to 100%) and still induce the firm to transfer its production facilities in order to satisfy the local content rule. However, a country can become too greedy and create a "boomerang" effect, whereby setting too high of a rate not only fails to attract investment but also causes the company to leave the country altogether (close its DC), see Munson and Rosenblatt [1997]. For the special case described here, Corollary 10.11 gives the local content percentage which "backfires" on country B.

Corollary 10.11 *For the case of identical plants and identical countries with country A having a lower tax rate than the other countries, the local content percentage Ω imposed by country B will result in the closing of all plants and DCs in that country if*

$$\Omega > \frac{P}{Ip+V}\left\{ V + \right.$$
$$\left. \frac{-W + \sum_{\omega=1}^{T} \ (1-t')\Delta CD + \frac{tW}{T} - (1-t)F - (t-t')D(P - Ip - C - V) \ /(1+\beta)^{\omega}}{\sum_{\omega=1}^{T}(t-t') \ NDp - (N-1)(1+\Delta)\vartheta CD - \vartheta CD - f - NDv - \frac{A}{T} \ /(1+\beta)^{\omega}} \right\}.$$

4. Illustrative Use of the Model: Effects of Financing, Tariffs, Local Content, Regional Trading Zones and Taxation

We explore the potential effects on a firm's location decisions from 1) financing incentives, 2) tax incentives, 3) regional trading blocks, and 4) local content requirements with the use of our model. For the base case problem we consider a company making a product which sells for $2,000 in all markets, irrespective of tariffs. The product has four major subassemblies with significant fixed and variable costs for the subassembly plants. Distribution centers, on the other hand, have lower

Market	Demand	Market	Demand
Europe		Asia	
United Kingdom	1,500	China	1,200
Ireland	300	India	1,200
Germany	3,500	Taiwan	300
France	1,500	Afghanistan	500
Romania	200	Pakistan	300
Sweden	800	Thailand	400
Portugal	400	Malaysia	500
Spain	600	Singapore	600
Norway	600	The Philippines	400
Finland	400	South Korea	1,500
Russia	1,000	Japan	3,000
The Netherlands	600		
Belgium	400	North America	
Luxembourg	50	Eastern U.S.	5,000
Denmark	800	Western U.S.	5,000
Switzerland	100	Mexico	600
Austria	300	Canada	1,500
Italy	1,000		
Greece	600		
Turkey	500		
Poland	500		
Hungary	400		
Bulgaria	200		
Ukraine	600		
Belarus	400		

Table 10.2. Demand in Each Market

fixed and variable costs. The firm sells to 40 countries in Europe, Asia, and North America. (The U.S. market is split into two "countries," East and West U.S., in order to account for large geographical distances. Of course, no trade tariffs are charged on flows between these two markets.) Table 10.2 shows the demand of each country.

The firm has identified 12 of the countries as potential locations for subassembly plants and distribution centers. Table 10.3 shows the relevant cost information. No financing is offered for the base case. Each country has a 40% effective income tax rate, and cash flows are discounted at 20% in each country. The time horizon is five years, and cost and demand factors do not change over time. All investments have an annual 20% depreciation rate. The subassembly transportation cost factor ϑ_i equals 20% for each subassembly i. The transportation costs for the final product are 2 cents per mile per unit. In addition, every final product that crosses a national border is charged a $200 per unit tariff ($40 per unit for each subassembly).

The transfer price p_{imkt} was set equal to a 20% markup over costs calculated as the sum of depreciation and fixed costs for plant i in country m divided by the total worldwide demand, the transportation cost

	U.K.	Ireland	Germany	France	Romania	Sweden
Subassembly 1						
Investment Cost	$500,000	$500,000	$500,000	$500,000	$12,000,000	$500,000
Annual Fixed Cost	$200,000	$200,000	$200,000	$200,000	$200,000	$200,000
Variable Cost	$350	$320	$350	$350	$140	$350
Subassembly 2						
Investment Cost	$500,000	$500,000	$500,000	$500,000	$250,000	$500,000
Annual Fixed Cost	$1.5 Mil.	$1.5 Mil.	$1.5 Mil.	$1.5 Mil.	$1.5 Mil.	$1.5 Mil.
Variable Cost	$300	$300	$300	$300	$300	$300
Subassembly 3						
Investment Cost	$1 Mil.	$1.2 Mil.	$1 Mil.	$1.2 Mil.	$1.5 Mil.	$1.2 Mil.
Annual Fixed Cost	$800,000	$800,000	$800,000	$800,000	$800,000	$800,000
Variable Cost	$300	$290	$300	$300	$270	$300
Subassembly 4						
Investment Cost	$200,000	$200,000	$200,000	$200,000	$200,000	$200,000
Annual Fixed Cost	$150,000	$150,000	$150,000	$150,000	$130,000	$150,000
Variable Cost	$380	$360	$380	$380	$350	$380
Distribution Center						
Investment Cost	$20,000	$20,000	$20,000	$20,000	$20,000	$20,000
Annual Fixed Cost	$12,000	$12,000	$12,000	$12,000	$12,000	$12,000
Variable Cost	$20	$20	$20	$20	$15	$20

	East U.S.	West U.S.	Mexico	China	India	Taiwan
Subassembly 1						
Investment Cost	$500,000	$500,000	$400,000	$400,000	$400,000	$400,000
Annual Fixed Cost	$200,000	$200,000	$200,000	$200,000	$200,000	$200,000
Variable Cost	$350	$350	$280	$280	$280	$300
Subassembly 2						
Investment Cost	$500,000	$500,000	$400,000	$400,000	$400,000	$400,000
Annual Fixed Cost	$1.5 Mil.	$1.5 Mil.	$1.5 Mil.	$1.5 Mil.	$1.1 Mil.	$1.5 Mil.
Variable Cost	$300	$300	$250	$250	$250	$260
Subassembly 3						
Investment Cost	$1 Mil.	$1 Mil.	$1.5 Mil.	$1.5 Mil.	$1.5 Mil.	$1.2 Mil.
Annual Fixed Cost	$800,000	$800,000	$800,000	$800,000	$800,000	$800,000
Variable Cost	$300	$300	$280	$280	$280	$280
Subassembly 4						
Investment Cost	$200,000	$200,000	$200,000	$200,000	$200,000	$200,000
Annual Fixed Cost	$150,000	$150,000	$130,000	$130,000	$130,000	$130,000
Variable Cost	$400	$390	$270	$280	$270	$290
Distribution Center						
Investment Cost	$20,000	$20,000	$20,000	$20,000	$20,000	$20,000
Annual Fixed Cost	$12,000	$12,000	$12,000	$12,000	$12,000	$12,000
Variable Cost	$20	$20	$15	$20	$15	$15

Table 10.3. Country-Specific Cost Data

of subassembly i from country m to the distribution center in country k and the unit variable production cost of i in m.

Base Case

The base case solution has a net present value of $38,940,519. India has plants for all four subassemblies. In addition, Mexico has a plant for subassembly 4. There are nine active distribution centers. They are located in all potential locations except for Ireland, Romania, and Taiwan. Each distribution center serves its local demand. The distribution center in India also serves the customers of all countries that do not have their own distribution center. The three potential DC locations that did not open were the ones with the smallest demands. As the reader can easily verify, the resulting facility configuration in this more complicated decision environment is along the same lines and fits nicely with the insights obtained from the special case discussion in Section 3.1.

Scenario 1 - Government Financing

In this scenario we examine whether government loans can induce investment in that country. In the base case, despite the very low variable cost, the high investment cost for subassembly 1 plant (mostly due to a lack of transportation, telecommunications and technology infrastructure) prevented its location in Romania. However, government financing can make production in Romania attractive. For this example, if Romania offers to finance all possible investments at an interest rate less than or equal to 33.2%, the plants for subassemblies 1 and 3 move from India to Romania (at higher interest rates the base case solution is still optimal).

The shift in locations to Romania has global distribution impacts. With the production network more spread out, it is optimal to now open distribution centers in all 12 potential locations. Also, the DC in India now only serves the markets in Asia which do not have their own DCs. The Romanian DC serves all other markets in Europe and North America without their own DCs.

Reducing the interest rate from 33.2% does not change the new allocation of demand, even when the rate is 0%. The NPV does increase, of course, as the interest rate decreases. Figure 10.1 shows the percent increase in NPV as a function of the interest rate offered by the Romanian government. At 0% the NPV is $44,349,287. The rate of 33.2% necessary to induce the configuration change is the one prescribed in Proposition 10.3 (i.e., $(\beta/1-t)$).

Scenario 2 - Tax Incentives

Ireland has traditionally offered low tax rates to try to increase foreign investment. Since Ireland has no locations in the base case, in Scenario

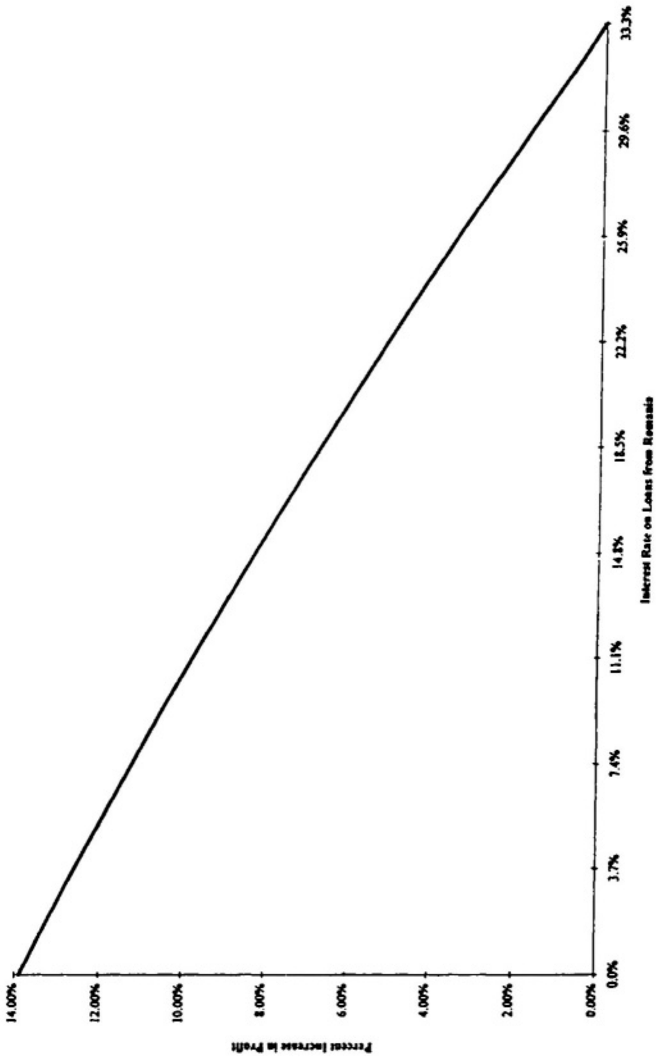


Figure 10.1. Scenario 1-Government Financing

2 we lower the tax rate in Ireland while keeping the rate in all other countries at 40%. It only takes a 4% reduction in tax rate to 36% to entice the firm to open a distribution center in Ireland to serve its own customers. All other demand allocations from the base case remain the same.

At 25%, the plant configuration as well as the DC configuration change. The plant for subassembly 1 moves from India to Romania, and the plant for subassembly 3 moves from India to Ireland. In addition, all 12 countries on the candidate location list open distribution centers. India continues to serve the customers in Asian markets without DCs, but all other markets without DCs are now served by Ireland. Another configuration change occurs at a 5% tax rate. The plant for subassembly 2 moves from India to Ireland. In addition, all DCs in Europe except for Ireland's close, and the Ireland DC now serves all markets which do not have their own DCs.

Figure 10.2 shows the percent increase in NPV as a function of Ireland's tax rate. The curve is piecewise-linear, with the kinks occurring at tax rates that cause a change in network configuration and/or demand distribution among DCs. At 0% the NPV is $45,073,946.

The results of our example for Scenario 2 clearly indicate that lower tax rates can encourage companies to change their location decisions, and the larger the difference between the rates of low-tax and high-tax countries, the more centralized the manufacturing/distribution network structure tends to become (i.e., similar insights to Proposition 10.6). In addition, firms can take advantage of tax rate differences by changing their intra-company transfer prices. (Overindulgence in such activity requires caution, however, because governments often monitor transfer prices.) We ran the model to take additional advantage of Ireland's lower tax rates by setting the transfer prices to cover the subassembly plants' fixed and variable costs only (assuming full satisfaction of worldwide demand). The DC and plant configurations were not always the same for the new transfer prices. For example, if Ireland's tax rate is set to 0%, Romania has the plant for subassembly 1, India has the plants for subassemblies 2 and 4, and Ireland has the plant for subassembly 3; and there are only two DCs – one in India which serves the Indian market only and one in Ireland which serves the rest of the world. The NPV increases substantially with the new transfer prices for lower tax rates and achieves its maximum difference (Figure 10.3) at the 0% tax rate, where the NPV is $50,306,318, a $5,232,372 increase compared to the old transfer price case.

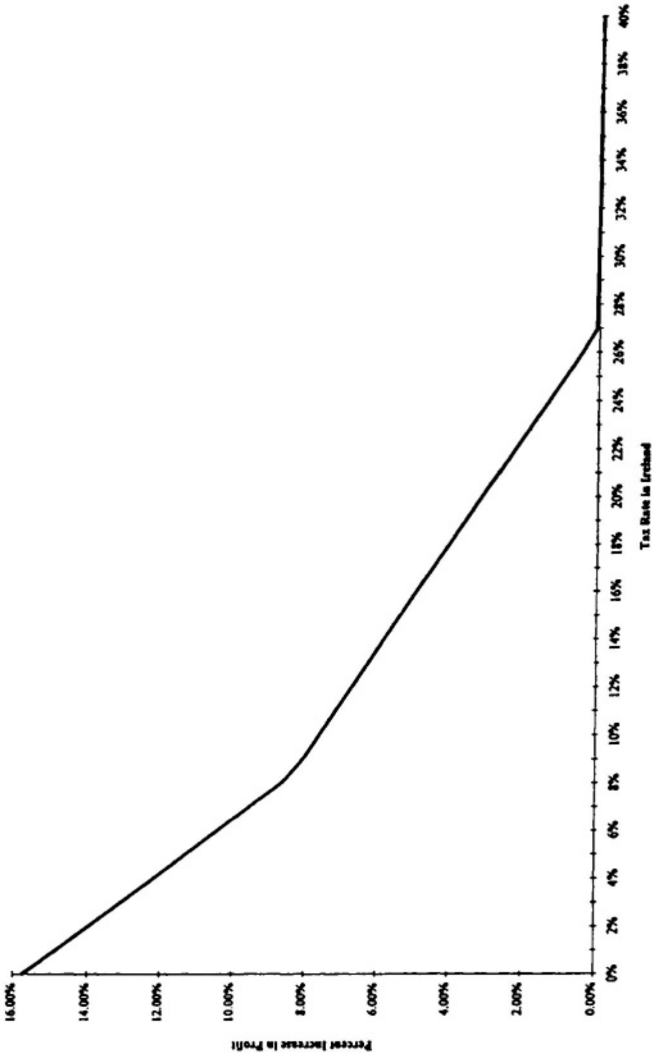


Figure 10.2. Scenario 2-Tax Incentives

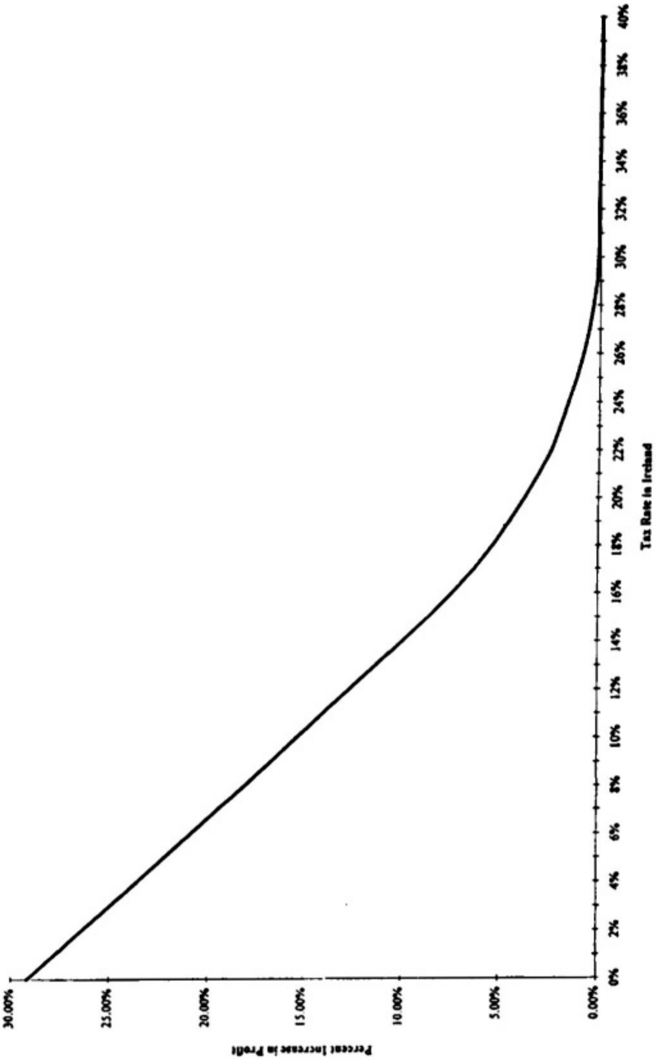


Figure 10.3. Scenario 2-Tax Incentives with Transfer Price Shift

Germany	Romania	Sweden	East U.S.	West U.S.	China	India	Taiwan
Germany	Romania	Sweden	East U.S.	West U.S.	China	India	Taiwan
U.K.	Norway		Mexico			Afghanistan	Thailand
Ireland	Finland		Canada			Pakistan	Malaysia
France	Russia						Singapore
Portugal	Switzerland						Philippines
Spain	Austria						So. Korea
Netherlands	Turkey						Japan
Belgium	Poland						
Luxembourg	Hungary						
Denmark	Bulgaria						
Italy	Ukraine						
Greece	Belarus						

Table 10.4. Demand Allocation to Distribution Centers for the Regional Trading Agreements Case

Scenario 3 - Regional Trading Agreements

Regional trading agreements such as NAFTA encourage trade among countries within the region. For this example, we ran the model eliminating the $200 tariff between countries who were members of the same regional trading blocks. We defined three trading blocks as 1) NAFTA: The U.S., Mexico, and Canada; 2) European Union: The U.K., Ireland, Germany, France, Portugal, Spain, The Netherlands, Belgium, Luxembourg, Denmark, Italy, and Greece; and 3) Asian Union: Taiwan, Thailand, Malaysia, Singapore, The Philippines, South Korea, and Japan.

Under the trading blocks scenario, the plants for subassemblies 1 and 3 move from India to Romania. Also, demand gets substantially reallocated among DCs to take advantage of the tariff-free trade zones. Table 10.4 shows the new demand allocation to DCs, which demonstrates the "regionalization" effect on the distribution network structure. One DC in each of the trading zones is supplying the demand for all countries within the free trading region (except for the DC in the West U.S. which was also open along with the East U.S. DC in the base case due to the significant, albeit tariff-free, distances). This result is along the same lines as the insights of Lemma 10.7 and Proposition 10.9. The NPV of Scenario 3 is $42,624,747, a $3,684,228 increase from the base case.

Scenario 4 - Local Content Requirements

We examine the effects of a local content rule which imposes the $200 tariff for domestic shipment from a DC unless the value of the subassemblies from domestic plants plus the value added at the DC is at least as high as the local content percentage times the final good's value.

Symbolically, we altered the program in the following way. If a country has a local content rule, then two new decision variables are added: 1) a binary variable Γ_{kt}, which is activated if country k has a DC in period t but the local content constraint is not satisfied; and 2) a nonnegative linear variable Λ_{kt} which represents the additional tariff cost for domestic shipment of goods that do not satisfy the local content requirement. Two new constraints per period are needed:

$$(1-\Omega)\, V_{kt} \sum_{j} Y_{kjt} \geq -M\Gamma_{kt} + \Omega \left(\sum_{i} \sum_{m \in L} p_{imkt} X_{imkt} \right), \; \forall\, k, t;$$

$$\text{and } \Lambda_{kt} \geq \Delta \left[\sum_{j \in L} Y_{kjt} - M\left(1 - \Gamma_{kt}\right) \right], \; \forall\, k, t\, (\Lambda_{kt} \geq 0),$$

where L is a set that contains all of the markets in the tariff-free zone, Ω is the local content percentage, M is a very large number, and Δ represents the additional tariff penalty per unit. The tariff penalty Λ_{kt} is then added to the DC transportation cost term a_{kt} in the objective function.

Trading blocks may impose local content requirements as well. For example, NAFTA uses a 62.5% rule in some cases. In the example problem, we examine the effect of a local content requirement in the Asian Union, where no subassembly plants exist in the optimal solution of Scenario 3. We obtained solutions for local content percentages ranging from 10% to 100%, incrementing by 10%.

At 10%, a plant for subassembly 4 is opened in Taiwan, and the distribution of demand among the DCs remains the same as in Scenario 3. At 30%, however, a major worldwide shift occurs. First, a plant for subassembly 3 is opened in Taiwan in addition to the subassembly 4 plant. Furthermore, the two plants in Romania are closed, and a plant for subassembly 1 is opened in both Mexico and India. This also causes a distribution center to be opened in Mexico to serve the domestic market there. Finally, the DC in Romania closes, and India now serves all of Romania's former markets.

Another change occurs at 50%. The plant configuration in Taiwan is the same, but the plants for subassembly 1 close in both India and Mexico (as does the DC in Mexico), and the subassembly 1 plant is reopened in Romania. This is the high investment, low variable cost plant in Romania. The transfer price per unit of subassembly 1 charged by Romania to Taiwan is \$319.98 vs. a price of \$404.57 charged when Taiwan "received" subassembly 1 from India. (Although the distance from Romania is further and the depreciation expense is larger than from a subassembly 1 plant in India, these effects on the transfer price are more than offset by the substantially lower variable production cost in Romania). So while in a physical sense the same subassemblies (3 and 4) are made in Taiwan under a 50% local content rule vs. a 30% rule, the total value of the product in Taiwan decreases since the transfer price of subassembly 1 decreases by \$84.59 compared to the price charged previously by the plant in India. In other words, Taiwan satisfies the local content rule not by increasing the numerator (value of local content), but by decreasing the denominator (value of the total product). This allows the DC in Taiwan to continue to satisfy local content with just two plants. Finally, the opening of the plant in Romania causes its DC to reopen, although India now continues to serve the markets of Norway, Finland, and Russia.

A third plant (for subassembly 1) opens in Taiwan under the 60% rule. Following the insights of Proposition 10.10, reasonable size increases in the local content rule of a trading region induce further manufacturing investment in the region. This move causes the subassembly 1 plant and the DC to close again in Romania. In addition, the subassembly 1 plants reopen in India and Mexico. The DC configuration returns to the 30% rule case.

Countries or trading blocks can become too greedy, however. An 80% rule in the Asian Union becomes too costly to try to accommodate. At that point all plants in Taiwan are closed. Now India has plants for all four subassemblies, and Mexico has plants for subassemblies 1 and 4. (The DC in Mexico remains open to serve the domestic market there.) Even the DC in Taiwan closes, so now India serves all of the Asian Union.

A similar "boomerang" effect occurred when we tried a local content rule for Germany and China only. The demand was insufficient in those countries to induce any plants to open. Instead, the DCs were forced to close in each case, "robbing" the local governments of any employment or income tax revenue. The insights of the "boomerang" effect of local content rule changes, as well as relevant conditions for it to occur, are also described in Corollary 10.11 of the previous section. Munson and Rosenblatt [1997] observe a similar effect for a different type of local content requirement.

Clearly local content rules can greatly affect the size and type of investment in a country or trading block, and the implications may affect the entire worldwide configuration. Companies should monitor such rules closely and take them into account early in the network design stages.

5. Conclusion

Government subsidies in facility financing, tariffs and regional trading rules, and favorable corporate taxation laws are main factors in explaining the structure of global manufacturing and distribution networks. We have presented a modeling framework that provides (for decision support purposes) the effects of such factors on the structure of global facility networks. From our analysis of special cases of the model, we obtained useful insights on the structure of such networks, which we present in a summary form below:

(i) when transportation of subassemblies is expensive, firms prefer centralized manufacturing and distribution structures;

(ii) increased trade tariffs favor gradual decentralization of the distribution networks;

(iii) differential tax rates tend to favor more centralized manufacturing and distribution network structures, with the low-tax countries attracting not only plants and DC investments but also serving a larger portion of the worldwide demand via their DCs;

(iv) formation of trading zones in certain environments (e.g., when the transportation cost of subassemblies is relatively insignificant compared to the final product) tends to have a "regionalization" effect on the facility network (i.e., the facility network becomes decentralized over trading zones but centralized within each zone);

(v) reasonable increases in local content requirements within a country or a trading zone entice firms to transfer facilities to that country or zone; however, inappropriate setting of local content requirements may result in exactly the opposite effect (i.e., withdrawal of all facilities from the country or zone);

(vi) even a small subsidy of facility financing by a country's government is in most cases adequate to attract facility investments in the country.

Acknowledgments

We wish to thank Charles Munson and Isaac Mizrahi for their contribution to this work.

Appendix

Proof of Proposition 10.1: The only component of the objective function affected by this decision is b_{nt}, i.e., the transportation cost of subassemblies to DCs. For each subassembly i, transportation costs are minimized by locating in the country containing the DC serving the largest demand. This way we minimize the additional trade tariffs (i.e., Δ component of the transportation cost). For the case where all countries have DCs in them, the subassembly plant location has no effect on the objective function, and thus the proposed solution of this proposition is among the optimal ones.

Proof of Proposition 10.2:

$$\text{NPV(1 DC)} - \text{NPV}(q \text{ DCs}) =$$
$$(q-1) \quad W + \sum_{\omega=1}^{T} \frac{(1-t)[\Delta CD(I\vartheta - 1) + F] - \frac{tW}{T}}{(1+\beta)^{\omega}} \Bigg).$$

The term inside the large parenthesis may be positive or negative. If it is positive, then only one DC should be used. If it is negative, then *N* DCs should be used because the difference NPV (IDC) – NPV (qDCs) is maximized when $q = N$. As $T \to \infty$, the term inside the large parenthesis equals $W + (1-t)\,[\Delta CD\,(I\vartheta - 1) + F]/\beta$, which is positive if $I\vartheta \geq 1$. If $I\vartheta < 1$, the expression is decreasing in *D* and equals zero when $D = [\beta W/(1-t) + F]/[\Delta C\,(1 - I\vartheta)]$.

Proof of Proposition 10.3: Due to the deductibility of interest, the cost of debt is $(1-t)\,r$. Since the firm's cost of capital is β, the company should accept a loan as long as the cost to service the debt is less than what it can earn on the principal, i.e., $(1-t)\,r < \beta$, or $r < \beta/(1-t)$. This is true for any size investment as long as the firm's debt ratio does not change significantly. Propositions 10.1 and 10.2 imply that, for the base case, all / plants will be located in a country with a DC. Since the countries are identical, the firm will increase NPV by placing all 7 plants and a DC in the country offering the financing, assuming that the interest rate satisfies the stated condition.

Proof of Lemma 10.4: Let country A be the one with the lower tax rate (t'). Country B will represent any of the other arbitrary countries with tax rate of $t\,(t' < t)$. We need to consider four cases:

Case Cl–all markets served from A and all plants are located in A;

Case C2–all markets served from B and all plants are located in B.

Case C3–all markets served from A and all plants are located in B;

Case C4–all markets served from B and all plants are located in A;

The cases only differ with respect to their annual cash flows (ACF). We use the notation ACF(Cl), ACF(C2), etc., to denote annual cash flows for cases Cl and C2, respectively. These expressions are:

$$\text{ACF(C1)} = (1-t')[NPD - (N + N\Delta - \Delta)CD - NI\vartheta CD - If - F - NIvD - NDV] + \frac{t'(IA+W)}{T};$$

$$\text{ACF(C2)} = (1-t)[NPD - (N + N\Delta - \Delta)CD - NI\vartheta CD - If - F - NIvD - NDV] + \frac{t(IA+W)}{T};$$

$$\text{ACF(C3)} = (1-t')[ND(P - Ip) - (N + N\Delta - \Delta)CD - F - NDV] + \frac{t'W}{T} + (1-t)[NDIp - (1+\Delta)NI\vartheta CD - If - NIvD] + \frac{tIA}{T};$$

$$\text{ACF(C4)} = (1-t)[ND(P - Ip) - (N + N\Delta - \Delta)CD - F - NDV] + \frac{tW}{T} + (1-t')[NDIp - (1+\Delta)NI\vartheta CD - If - NIvD] + \frac{t'IA}{T}.$$

After subtracting the annual cash flows and rearranging terms:

$$\text{ACF(C1)} - \text{ACF(C2)} = (t-t')[NPD - (N + N\Delta - \Delta)CD - NI\vartheta CD - If - F - NIvD - NDV - \frac{IA+W}{T}];$$

$\text{ACF(C1) - ACF(C3)} = (t-t')[NDIp - NI\vartheta CD - If - NIvD - \frac{IA}{T}] + (1-t)\Delta NI\vartheta CD;$

$\text{ACF(C1) - ACF(C4)} = (t-t')[ND(P-Ip) - (N - N\Delta - \Delta)CD - F - NDV - \frac{W}{T}] + (1-t')\Delta NI\vartheta CD.$

All three differences are positive due to the assumption of the price structure ensuring nonnegative profits in each country of operation. Thus, Case 1 has the largest NPV, which proves the lemma.

Proof of Lemma 10.5: We need to consider five cases:

Case C5 – all markets without a DC are served from A and all plants are located in A;

Case C6 – all markets without a DC are served from B, A has a DC, and all plants in B;

Case C7 – all markets without a DC are served from A and all plants are located in B;

Case C8 – all markets without a DC are served from B, A has a DC, and all plants in A;

Case C9 – country A has no facilities.

The cases only differ with respect to their annual cash flows. These expressions for C5-C8 are:

$\text{ACF(C5)} = (1-t')\{[N-(q-1)]D(P-Ip) - CD - (N-q)(1+\Delta)CD + NDIp - I\vartheta CD - (q-1)I\vartheta CD(1+\Delta) - (N-q)I\vartheta CD - If - F - NIvD - [N-(q-1)]DV\} + \frac{t'(IA+W)}{T} + (q-1)\{(1-t)[D(P-Ip) - CD - F - DV] + \frac{TW}{T}\};$

$\text{ACF(C6)} = (1-t)\{(N-1)D(P-Ip) - (q-1)CD - [N-(q-1)](1+\Delta)CD + NDIp - I\vartheta CD - (q-1)I\vartheta CD(1+\Delta) - (N-q)I\vartheta CD - If - (q-1)F - NIvD - (N-1)DV\} + \frac{t(IA+(q-1)W)}{T} + (1-t')[D(P-Ip) - CD - F - DV] + \frac{t'W}{T};$

$\text{ACF(C7)} = (1-t')\{[N-(q-1)]D(P-Ip) - CD - (N-q)(1+\Delta)CD - F - [N-(q-1)]DV\} + \frac{t'W}{T} + (1-t)\{NDIp - I\vartheta CD - (N-1)I\vartheta CD(1+\Delta) - If - NIvD\} + \frac{tIA}{T} + (q-1)\{(1-t)[D(P-Ip) - CD - F - DV] + \frac{tW}{T}\};$

$\text{ACF(C8)} = (1-t)\{(N-1)D(P-Ip) - (q-1)CD - [N-(q-1)](1+\Delta)CD - (q-1)F - (N-1)DV\} + \frac{t(q-1)W}{T} + (1-t')\{NDIp - I\vartheta CD - (N-1)I\vartheta CD(1+\Delta) - If - NIvD\} + \frac{t'IA}{T}\ (1-t')[D(P-Ip) - CD - F - DV] + \frac{t'W}{T}.$

After subtracting the annual cash flows and rearranging terms:

$\text{ACF(C5) - ACF(C6)} = (t-t')\{NDIp - I\vartheta CD - (q-1)I\vartheta CD(1+\Delta) - (N-q)I\vartheta CD - If - NIvD - \frac{IA}{T} + (N-q)D[(P-Ip) - V - (1+\Delta)C]\} + (1-t)(1+\Delta)CD;$

$\text{ACF(C5) - ACF(C7)} = (t-t')\{NDIp - NI\vartheta CD - If - NIvD - \frac{IA}{T}\} + I\vartheta CD\Delta[(1-t)(N-1) - (1-t')(q-1)];$

$\text{ACF(C5) - ACF(C8)} = (t-t')\{(N-q)D[(P-Ip) - V - (1+\Delta)C]\} + (1-t')(N-q)I\vartheta CD\Delta + (1-t)(1+\Delta)CD$

All three differences are positive due to the assumption of the price structure ensuring nonnegative profits in each country. Finally, Case C5 dominates Case C9 using the following argument. Suppose that (q-1) DCs have already been located in the higher tax rate countries and they serve their home market only. The final DC which serves the rest of the markets must be located in either country A or country B. This is a one-DC location problem described by Lemma 10.4, which implies that the DC should be located in country A.

Proof of Proposition 10.6: The first statement is a direct result of Lemmas 10.4 and 10.5. The two net present values, then, that need to be considered are:

$$\text{NPV(1 DC)} = -IA - W + \sum_{\omega=1}^{T} \{(1-t')[NPD - (N + N\Delta - \Delta)CD - NI\vartheta CD - If - F - NIvD - NDV] + \tfrac{t'(IA+W)}{T}\}/(1+\beta)^{\omega};$$

$$\text{NPV}(q \text{ DCs}) = -IA - qW + \sum_{\omega=1}^{T} ((1-t')\{[N-(q-1)]D(P-Ip) - CD - (N-q)(1+\Delta)CD + NDIp - I\vartheta CD - (q-1)I\vartheta CD(1+\Delta) - (N-q)I\vartheta CD - If - F - NIvD - [N-(q-1)]DV] + \tfrac{t'(IA+W)}{T}\} + (q-1)\{(1-t)[D(P-Ip) - CD - F - DV] + \tfrac{tW}{T}\})/(1+\beta)^{\omega}.$$

Subtracting the two:

$$\text{NPV(1 DC) - NPV}(q \text{ DCs}) = (q-1)(W + \sum_{\omega=1}^{T} \{(1-t')\Delta CD(I\vartheta - 1) + (1-t)F - \tfrac{tW}{T} + (t-t')D[(P-Ip) - C - V]\}/(1+\beta)^{\omega}).$$

As with the base case, if the term inside the large parenthesis is positive, then only one DC should be used (in country A). If it is negative, then *N* DCs should be used because the difference is maximized when $q = N$.

Proof of Lemma 10.7: Each DC that is built requires an investment plus an annual fixed cost. The only benefit of building more than one DC is to avoid the tradetariff Δ. The proof follows since transportation within a tariff-free zone has no Δ component.

Proof of Lemma 10.8: Along the lines of Proposition 10.2, it can be shown that either one DC should be built for the whole company or one DC should be built in each zone. If *Z* DCs are built, then subassembly transportation costs are minimized by building all *I* plants somewhere in zone *Z*1. If one DC is built, then it should be built in $Z1$to minimize the transportation costs of the final product. In that case as well, subassembly transportation costs are minimized by building those plants in Zl.

Proof of Proposition 10.9: Let z' countries be located in the zone *Z*1 that contains the most countries. Then by Lemmas 10.7 and 10.8, the NPVs of the two possible configurations are:

$$\text{NPV(1 DC)} = -IA - W + \sum_{\omega=1}^{T} \{(1-t)[NPD - z'CD - (N-z')(1+\Delta)CD - NI\vartheta CD - If - F - NIvD - NDV] + \tfrac{t(IA+W)}{T}\}/(1+\beta)^{\omega};$$

$$\text{NPV}(Z \text{ DCs}) = -IA - ZW + \sum_{\omega=1}^{T} \{(1-t)[NPD - NCD - z'I\vartheta CD - (N-z')I\vartheta CD(1+\Delta) - If - ZF - NIvD - NDV] + \tfrac{t(IA+ZW)}{T}\}/(1+\beta)^{\omega}.$$

$$\text{NPV(1 DC)} - NPV(Z\ DCs) = (Z-1)(W + \sum_{\omega=1}^{T} \{(1-t)[\tfrac{N-z'}{Z-1}\Delta CD(I\vartheta - 1) + F] - \tfrac{tW}{T}\}/(1+\beta)^{\omega}).$$

As with the base case, a centralized distribution network is optimal if and only if the term inside the large parenthesis is positive (it is one of alternate optima if the term equals zero). That term only differs from the corresponding base case by the factor $(N - z')/(Z - 1)$, which is greater than 1 assuming that at least one zone other than Z\ contains more than one country. Thus, the term $\Delta CD\,(I\vartheta - 1)$ is weighted more heavily than under the base case. So when that term is positive $(I\vartheta > 1)$, the entire term inside the large parenthesis is more likely to be positive than under the base case, but when that term is negative $(I\vartheta < 1)$, the entire term inside the large parenthesis is more likely to be negative than under the base case.

Proof of Proposition 10.10: The denominator represents the NPV of the cost of moving one subassembly plant from country A to country B. The numerator is the NPV of the cost of not satisfying the local content rule. If the local content rule is not satisfied, then with a tariff of Δ charged to domestic shipments of the final product, the firm would desire to close the DC in country B and serve that market from the DC in country A (due to the lower tax rate and elimination of the tariff-free domestic shipment). If this occurs, then the firm will have one less distribution center but will incur a tariff on final products shipped to country B. The numerator must be positive or it would not have been optimal to have a DC in country B in the first place.

Proof of Corollary 10.11: Let Φ' be the largest Φ which satisfies (10.1). Then the local content percentage is too high if $\Phi^* > \Phi'$, which occurs if $[\Omega I p - (1 - \Omega)\,V]/p$ is greater than the right-hand-side of (10.1).

References

Arnzten, B.C.; Brown, G.G.; Harrison, T.P.; and Trafton, L.L. (1995), "Global Supply Chain Management at Digital Equipment Corporation," *Interfaces,* 69-93.

Bartmess, A. and Cerny, K. (1993), "Building Competitive Advantage Through a Global Network of Capabilities," *California Management Review,* 78-103.

Breitman, R.L. and Lucas, J.M. (1987), "PLANETS: A Modeling System for Business Planning," *Interfaces,* 17, 94-106.

Cohen, M.A.; Fisher, M.L.; and Jaikumar, R. (1989), "International Manufacturing and Distribution Networks: A Normative Model Framework," *Managing International Manufacturing,* Kasra Ferdows (Ed.), North Holland, 67-93.

Cohen, M.A. and Lee, H.L. (1988), "Strategic Analysis of Integrated Production-Distribution Systems: Models and Methods," *Operations Research,* 36, 2, 216-228.

Cohen, M.A. and Lee, H.L. (1989), "Resource Deployment Analysis of Global Manufacturing and Distribution Networks," *Journal of Manufacturing and Operations Management,* 81-104.

Davis, T. (1993), "Effective Supply Chain Management," *Sloan Management Review,* 35-45.

Huchzermeier, A. and Cohen, M.A. (1996), "Valuing Operational Flexibility Under Exchange Rate Risk," *Operations Research,* 44, 1, 100-113.

Hodder, J.E. and Dincer, M.C. (1986), "A Multifactor Model for International Plant Location and Financing Under Uncertainty," *Computers and Operations Research,* 13, 5, 601-609.

Hodder, J.E. and Jucker, J.V. (1985a), "A Simple Plant-Location Model for Quantity Setting Firms Subject to Price Uncertainty," *European Journal of Operational Research,* 21, 39-46.

Hodder, J.E. and Jucker, J.V. (1985b), "International Plant Location Under Price and Exchange Rate Uncertainty," *Engineering and Production Economics,* 9, 225-229.

Jucker, J.V. and Carlson, R.C. (1976), "The Simple Plant Location Problems Under Uncertainty," *Operations Research,* 24, 6, 1045-1055.

MacCormack, A.D.; Newman, L.J.; and Rosenfield, D. (1994), "The New Dynamics of Global Manufacturing Site Location," *Sloan Management Review,* Summer, 69-80.

Munson, C.L. and Rosenblatt, M.J. (1997), "The Impact of Local Content Rules on Global Sourcing Decisions," Special Issue on Global Supply Chain Management in *Production and Operations Management,* forthcoming.

Revelle, C.S. and Laporte, G. (1996), "The Plant Location Problem: New Models and Research Projects," *Operations Research,* 44, 6, 864-874.

Chapter 11

MANUFACTURING PLANNING OVER ALTERNATIVE FACILITIES: MODELING, ANALYSIS AND ALGORITHMS

S. David Wu, Hakan Golbasi

Manufacturing Logistics Institute

Department of Industrial and Systems Engineering

Lehigh University

Bethlehem, PA 18015

david.wu@lehigh.edu

Abstract We propose a planning model for multiple products manufactured across multiple manufacturing facilities sharing similar production capabilities. The need for cross-facility capacity management is most evident in high-tech industries that have capital-intensive equipment and a short technology life cycle. Our model is based on an emerging practice in these industries where product managers from business units dictate manufacturing planning in facilities that are equipped to produce their products. We propose a multicommodity flow network model where each commodity represents a product and the network structure represents linked manufacturing facilities capable of producing the products. We analyze in depth the product-level (single-commodity, multi-facility) subproblem when the capacity constraints are relaxed. We prove that even the general-cost version of this uncapacitated subproblem is NP-complete. We develop a shortest-path algorithm for this problem and show that it achieves optimality under special cost structures. We analyze and pinpoint specific cases where the algorithm fails to produce optimal solutions. To solve the overall (multicommodity) planning problem we develop a Lagrangian decomposition scheme, which separates the planning decisions into a number of *single-product, multi-facility subproblems and a resource subproblem.* Through extensive computational testing, we demonstrate that the shortest path algorithm is an effective heuristic for the MIP subproblem, yielding high quality solutions with only a fraction (roughly 2%) of the computer time.

J. Geunes et al. (eds.),
Supply Chain Management: Models, Applications, and Research Directions, 279–316.

1. Introduction

This research is motivated by production problems in the electronics, semiconductor, and telecommunications industries. These industries struggle with their production planning problems in an increasingly complex and rapidly changing supply chain environment. Specifically, to better utilize their capital-intensive equipment they are pressured to produce a wide variety of products in each of their production facilities. However, since these products may each belong to a different supply channel operating under different delivery and outsourcing contracts and demand characteristics, production planning decisions are often relegated to *product managers* who are most familiar with their specific customer and supplier issues. On the other hand, *production managers* must consider resource consolidation and capacity management issues in a consistent fashion across manufacturing facilities. These competing viewpoints complicate supply chain planning significantly.

Coordinating production under a complex supply structure is not a new problem. However, two recent trends in these industries exacerbate the intensity of the problem. First, the trend toward increased market responsiveness intensifies the operational dependency within the supply chain. In the past, excess inventory was generally used to reduce the impact of variation across different facilities. Today, most manufacturers are moving away from carrying substantial inventories. Second, the rate of technological innovation significantly shortens the life span of manufacturing equipment, which in turn increases the cost of manufacturing capacity. This combined with increased product variety and decreased product volumes prompts manufactures to cross-load their manufacturing facilities.

In this paper, we focus on cross-facility operational planning faced by high-tech manufacturing companies in a supply chain environment. Our research is motivated by experiences in a production management system at a major semiconductor manufacturer for their world-wide supply base, and by the quantitative supply chain literature. Quantitative analysis of supply chain management has been focusing on channel design in general and stocking policies in specific using extensions of inventory, game-theoretic, and strategic models (Tayur et al., 1999). Cohen and Lee (1988), (1989), Sterman (1989) and Davis (1993) are among the pioneers who made significant early contributions. Various developments of these models remain an area of active research (c.f., Tayur et al. 1999, Lee et al. 1995, Hahm and Yano 1995a,b and Arntzen et al. 1995). In addition to supply chain design, coordinating various aspects of supply chain operations has been an area of active research as well. This line of work is exemplified by Vidal and Goetschalckx (1997), Hahm and Yano (1995a,b), and Ertogral and Wu (1999). A related, but distinctively different, line of research focuses on the extension of production models in the context of MRP systems (c.f., Billington et al. 1983, Carlson and Yano 1983, Gupta and Brennan 1995, Balakrishnan and Geunes 2000). The focus here is manufacturing planning in the context of multi-echelon and multi-facility production. This line of research is rooted in multi-level, multi-period, capacitated lot-sizing models. A number of surveys (c.f., Bahl et al. 1987, Goyal and Gunasekaran 1990, Baker 1993, Kimms 1997) provide an excellent overview for research in this area. While our proposed model can be linked directly to the multi-level lot sizing literature, it has two distinctive features that were not previously addressed: first is the explicit consideration of *facility selection* decisions. Most existing work assumes either a single facility, or multiple tiers of facilities as defined by the product structure, but the *facility selection* decisions are given *a priori.* Second, we study a single-item subproblem (with facility selection decisions) that has been overlooked in the literature. Unlike it's single-facility counterpart, this subproblem is NP-complete even in the uncapacitated case. Despite this, we show analytically and empirically that a shortest path algorithm can be extremely effective.

2. A Multi-Facility Production Model

We now consider a multi-facility production model where a set of end-items is to be produced in multiple facilities over multiple periods. Each end-item has a bill of materials described by a *product structure.* In addition, there is a *supply structure* where a set of alternative facilities could be setup to produce each item described in the product structure. Figure 11.1 illustrates the product and the supply structure. The *product structure* in Figure 11.1 can be represented by an $n \times n$ matrix $[a_{ik}]$ where a_{ik} is the number of units of item i (directly) needed to produce one unit of item k. We define a *supply structure matrix* $[r_{ij}]$ where $r_{ij} = 1$ if facility j could be used to produce item i, and $r_{ij} = 0$ otherwise. In this paper, we will focus our attention on the supply structure.

2.1 A Multicommodity Flow Model

Production planning in the above multi-facility environment is complex in that the facility selection decisions are combined with multi-stage, multi-item, multi-period production decisions. To approach this problem we take the viewpoint of a subset of manufacturing facilities in the supply network. Each manufacturing facility can produce a variety of products (items) $(i = 1, 2, \ldots, n)$ over multiple periods $(t = 1, 2, \ldots, T)$, while

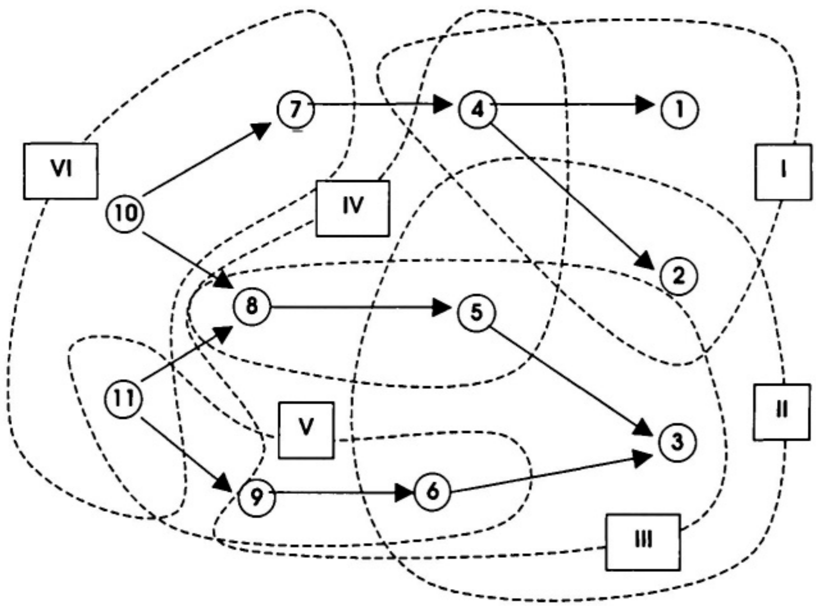


Figure 11.1. A Supply Network with Three End-Items (1, 2, and 3), Six Facilities (I to VI) and a Maximum of Three Alternative Facilities for an Item

each item i can be produced in a specified set of alternative facilities $(j \in J_i, J_i \subseteq 1, 2, \ldots, m)$. Each facility can be setup to perform certain production processes with a setup cost. Now consider a multicommodity network G(N,A) where each item i corresponds to a commodity in the network. Let D_t^i denote internal demands for item i in period t as defined by the end-item demand and the product structure. Suppose D_t^i can be determined *a priori*; we can then define a multicommodity flow network corresponding to the supply network as shown in Figure 11.2. This multicommodity flow network has three main parts: a set of source nodes representing demand dispatching points, a set of sink nodes representing demand fulfillment points, and a set of *production subnets* in between, each representing multi-period production to be carried out by a facility. Each commodity (product) i has a source node s^i, and T sink nodes d_t^i, one for each period t.

We first describe the overall network structure. The input flow for source node s^i is the total demand over T periods for item i $(\sum_{t=1}^{4} D_t^i)$, and the outflow on sink node d_t^i is the demand for item i in period t, (D_t^i). The presence of an arc going from source node s^i to facility subnet j signifies the fact that facility j can be setup to produce item

i. These arcs are specified by the *supply structure matrix,* i.e., there is an arc (i, j) corresponding to each non-zero entry of matrix $[r_{ij}]$. The subnets between the set of source and sink nodes represent production facilities shared by the products. The arcs going from facility subnets $j = 1, 2, \ldots, J_i$, period t, to sink node d_t^i represent the requirement that item i's demand in period t is to be fulfilled by the production and/or inventory from all (or some subset) of its J_i alternative facilities. Note that the *production subnet* can be further "customized" according the structure of each manufacturing facility. For example, the structure in Figure 11.2 shows the familiar single-stage multi-period lot-sizing model. This can be extended to a multi-stage model (c.f., Afentakis 1984), or other multi-period models. To streamline the analysis, we assume single-stage facility subnets throughout the paper.

We now characterize arc labels in the multicommodity network. Each arc in the network is characterized by (f^i, c^i, u): arc flow f^i, per unit cost c^i, and arc capacity u. The interpretation of these values varies according to the types of arcs. The arcs going from source nodes s^i to the facility subnets j's are *facility selection arcs* $A_s \subset A$, characterized by (x_j^i, c_j^i, u_{ij}) : x_j^i represents the total production of item i to be performed at facility j over $t = 1, \ldots, T$, c_j^i represents the (per unit) costs that differentiate the facilities (e.g., by quality history, reputation), and capacity u_{ij} represents the maximum amount of item i that can be produced in facility j. When the dynamic lot-sizing model is used in the *production subnet,* two types of arcs are used: arcs going from left to right are *production arcs* $A_p \subset A$, characterized by $(x_{jt}^i, c_{jt}^i, cap_{jt})$: production quantity x_{jt}^i, unit production cost c_{jt}^i, and production capacity cap_{jt}. Arcs going from the top down are *inventory arcs* $A_I \subset A$ characterized by $(I_{jt}^i, h_{jt}^i, inv_{jt})$: inventory carried from period t to $t+1$, I_{jt}^i, unit inventory holding cost h_{jt}^i, and inventory limit inv_{jt}. Finally, an arc going from facility j in period t to sink node d_t^i belongs to the *demand arcs* $A_d \subset A$, characterized by $(b_{jt}^i, r_{jt}^i, cap_{jt})$: b_{jt}^i represents facility j's contribution to demand D_t^i, r_{jt}^i represents the per unit transportation cost from facility j to the demand point i, and cap_{jt} is the transportation capacity in period t.

Given the above specification, we can define the *general multicommodity constraints* (11.1) and (11.2) as follows:

General arc capacity constraints for all arcs

$$x_j^i \le u_{ij}, \ \forall (i, j) \in A_s,$$

$$\sum_{i=1}^{n} \beta_{jt}^i \le f_{jt}^i u_{jt}, \ \forall (j, t) \in A_p \cup A_I \cup A_d, \tag{11.1}$$

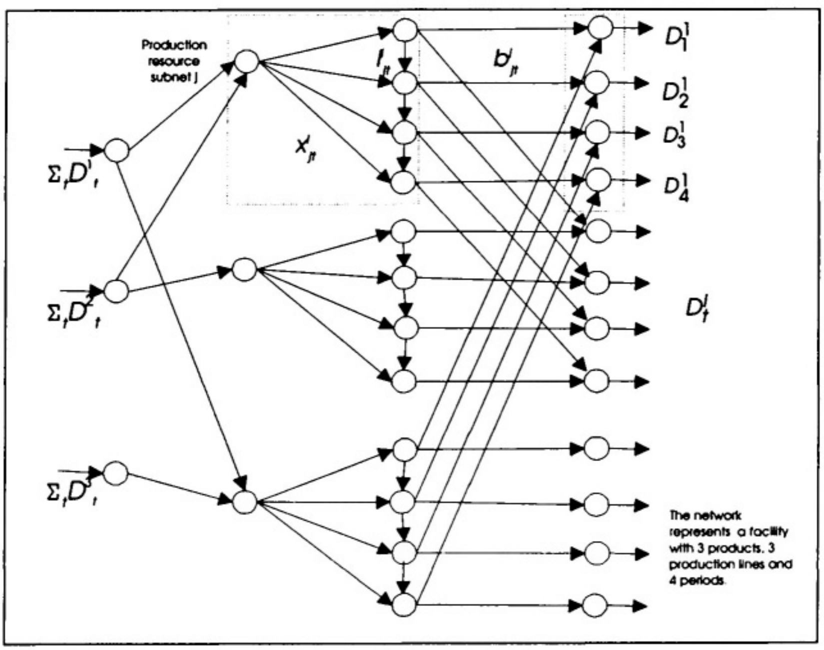


Figure 11.2. A Multicommodity Flow Network for the Production Problem

where β is the capacity consumption rate. Denote α as the node-arc incidence matrix for the multicommodity flow network G(N,A) and e^i the net balance flows for commodity i. The mass balance constraints are as follows:

Mass balance constraints for each commodity

$$\alpha f^i = e^i, \ \forall i \in N. \tag{11.2}$$

In addition to the general multicommodity flow constraints as specified by the network structure G(N,A), we define additional constraints for each facility submodel. Consider the multi-period lot-sizing model we use for all facility subnets. While the inventory balance constraints are part of the mass balance constraints (11.2), we need to define additional constraints due to setups. Let α^i_{jt} denote the rate of capacity consumption for setup activities; δ^i_{jt} is a binary variable indicating the existence of a setup for item $i(= 1, \ldots, n)$ at facility $j(= 1, \ldots, m)$ in period $t(= 1, \ldots, T)$. The *production specific constraints* are as follows:

Production capacity constraints for production and setup

$$\sum_{i=1}^{n} (\beta^i_{jt} x^i_{jt} + \alpha^i_{jt} \delta^i_{jt}) \leq cap_{jt}, \ \forall (j,t) \in A_p. \tag{11.3}$$

Setup constraints

$$x^i_{jt} \leq M \delta^i_{jt}, \ \forall i \in N, (j,t) \in A_p, \tag{11.4}$$

where *M* is a sufficiently large constant.

Demand for each item must be satisfied

$$\sum_{j=1}^{m} b^i_{jt} = D^i_t, \ \forall t, i \in N_0. \tag{11.5}$$

The end-item demand triggers internal demands in the supply chain as defined by the product structure.

Denoting K^i_j as the set-up cost for item i at facility j, we state an objective function with all cost components defined for the *production subnet.* A cost component unique to each product i, say $S_i(f^i, c^i, u)$, could be added to the objective to reflect special requirements imposed by the supply channel of i. Since this does not affect our analytic results significantly, we will not consider this cost term here. Thus, a multicommodity flow formulation of the multi-facility production problem (P) is as follows:

Problem (P):

$$Minimize\ z = \sum_{t=1}^{T}\sum_{i=1}^{n}\sum_{j=1}^{J_i}(c_{jt}^{i}x_{jt}^{i} + K_{j}^{i}\delta_{jt}^{i} + h_{jt}^{i}I_{jt}^{i} + r_{jt}^{i}b_{jt}^{i})$$

s.t.

<*General Multicommodity Flow Constraints* (11.1)-(11.2)>
<*Production Specific Constraints* (11.3)-(11.5)>
<*Nonnegativity Constraints*> $f^i, x_{jt}^i, I_{jt}^i, b_{jt}^i \geq 0$
<*Binary Constraints*> $\delta_{jt}^i \in \{0,1\}$

It is useful to note that in this multicommodity flow model, only the capacity constraints (11.1) and (11.3) are bundling constraints. All the other constraints can be decomposed by commodity (product). (P) is a multi-period, multi-item, multi-facility production planning model.

3. Model Analysis

To explore special subproblem structures that will later help the solution of model (P), we consider two submodels in the following sections.

3.1 Uncapacitated Single-Item, Multi-Facility Model without Transportation Costs

As stated above, model (P) can be decomposed by commodity after relaxing the bundling capacity constraints (11.1) and (11.3). We first consider a capacity-relaxed single-item, multiple-facility subproblem (without transportation costs) for commodity i as follows:
(P_i)

$$Minimize \sum_{t=1}^{T}\sum_{j=1}^{J_i}(c_{jt}x_{jt} + K_j\delta_j + h_{jt}I_{jt})$$

s.t.

<*mass balance constraints for commodity* i (11.2)>
<*setup constraint for commodity* i (11.4)>
<*nonnegativity and integrality constraints for commodity* i >

Subproblem (P_i) represents a subset of the decision problem for the product manager of i who must decide where to produce her product among a certified[1] set of manufacturing facilities J_i. Note that we dropped the cost component $S_i(f^i, c^i, u)$ from the objective, hence the facility selection and the transportation costs are assumed to be zero. Figure 11.3 depicts a subnetwork defined by commodity 1 (for problem (P_1)) corresponding to the example in Figure 11.2.

While the above facility-selection/lot-sizing problem has not been well studied, the *single*-facility, single-item, uncapacitated lot-sizing problem has been studied intensively in the literature. Despite the binary variables it is well known that this problem can be solved in polynomial time using Wagner-Whitin (1958) type algorithms. In recent years, more efficient implementations of Wagner-Whitin algorithms have been developed (c.f., Federgruen and Tzur 1991, and Wagelmans et al. 1992), which have complexity of $O(n \log n)$ or better. Embedding these polynomially solvable problems as submodels, the *multi-item, capacitated* lot-sizing problems are frequently solvable in a reasonable amount of time for realistic size problems (cf. Tempelmeier and Derstroff 1996). Since a primary new consideration in our model (P) is the selection of alternative facilities, it is important to know the structure and the complexity of subproblem (P_i). This analysis is important to the solution of (P) since (1) (P_i) has the form of a mixed integer program, and (2) there is no straightforward (efficient) decomposition from the multi-facility case to a single-facility problem. In the following, we first show that (P_i) is *NP-complete* when the holding cost h_{jt} is general and not restricted in sign. We then show that an efficient algorithm exists under a special set of conditions. Later in Section 4 we show computationally that under general cost conditions this algorithm is an effective heuristic, solving most instances of (P_i) optimally.

Theorem 11.1 *[Non-splitting property] There exists an optimal solution to the uncapacitated, single-item, multiple facility problem (P_i) such that item i's demand in period t is satisfied by the production or the inventory of exactly one of the J_i facilities, i.e., exactly one b^i_{jt} among $j \in J_i$ is non-zero $(= D^i_t)$ for each period t.*

Proof: See Appendix.

As we shall see in the following exploration, the insight provided by the *non-splitting property* plays an important role in the development

[1]The notion of facility certification is important in Semiconductor Manufacturing where a product can only be produced in a facility that has been pre-certified for quality and yield.

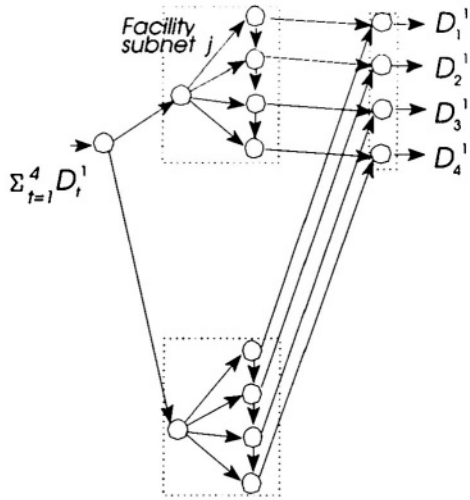


Figure 11.3. Subnetwork Corresponding to Commodity 1, with 2 alternative facilities, 4 periods

of solution algorithms for the uncapacitated subproblem (P_i) as well as the overall problem (P). Nonetheless, despite its promising outlook, subproblem (P_i) can be only solved efficiently under more restrictive conditions due to the existence of several special cases. In the following, we first show that under generalized cost conditions subproblem (P_i) is *NP-complete.*

Theorem 11.2 *The uncapacitated, single-item, multiple-facility problem is NP-complete when the inventory holding cost h_{jt} is a generalized cost coefficient not restricted in sign.*

Proof: See Appendix.

Upon examining the proof it should be clear why *NP-completeness* is only constructed for the case where arc cost is not restricted in sign. In the following, we show that the *NP-complete* status remains when the demands or the setup costs are constant.

Corollary 11.3 *The problem stated in Theorem 11.1 remains NP-complete when the period demands D_t are constant over periods $t = 1, \ldots, T$.*

Corollary 11.4 *The problem stated in Theorem 11.1 remains NP-complete when the setup costs K_j are constant across facilities $j = 1, \ldots, m$.*

In the following, we go on to show that despite the *NP-complete* status of problem (P_i) and its variations, under special conditions a shortest path algorithm solves problem (P_i).

Proposition 11.5 *The uncapacitated, single-item, multiple-facility problem (P_i) can be solved in polynomial time using a shortest path algorithm if the following conditions hold:*

(i) No simultaneous production of item i over more than one facility can take place in a given period. In other words, $x^i_{jt}x^i_{kt} = 0 \, \forall i, j, k \neq j, t$.

(ii) No production of item i will be scheduled at all if there is inventory carried over from a previous period in one of the facilities. In other words, $x^i_{jt}I^i_{kt-1} = 0 \, \forall i, j, k, t$.

Proof: We will state the proof using a familiar graphical representation as in Figure 11.4. The first row of nodes denotes facility 1 and the i^{th} row denotes facility J_i. There are T + 1 time epochs: 0, 1, 2, 3, and a period is the interval between epochs, i.e., between epochs 0 and 1 is period $t = 1$, and between 1 and 2 is period $t = 2$, etc. A horizontal arc denotes the production that satisfies all the periods' demand within the time epochs. The arc cost includes production, inventory and setup costs. A vertical arc denotes a switch from one facility to another and the cost associated with these arcs is 0. There is an artificial source and sink, and arcs adjacent to these nodes have 0 cost. In a general graph for each time epoch there are arcs for all the facility periods. So production of an item may switch from one facility to any other in different periods. From condition (ii), there will be production scheduled for item i only if there is no inventory carried over from a previous period in one of the facilities. In other words, in an optimal solution exactly one arc will be chosen to enter a given node in the network. On the other hand, condition (i) states that there can be no simultaneous production of items in more than one facility in any given period. In other words, in an optimal solution exactly one arc will be chosen to leave a node in the network. This means that an optimal production schedule corresponds to a source-to-sink path in the network, and it corresponds to a shortest cost path. $\diamond$

Unfortunately, there are cases where conditions (i) and (ii) in Proposition 11.5 do not hold in the optimal solution. This is caused by special cases such as the following: under completely general production and inventory holding costs, it could be optimal to produce x^i_{jt} in a facility j in period t, and hold this amount in inventory for a future period $t + l$ $(l > 1)$ without using it in periods $t + 1, \ldots, t + l - 1$. Since more than

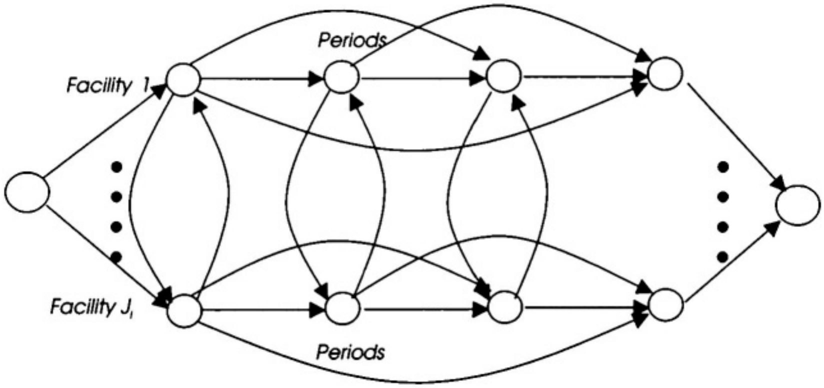


Figure 11.4. The Multi-Facility Production Problem as a Shortest Path Problem

one facility may produce, the demands in periods $t+1,\ldots,t+l-1$ could be satisfied by a number of facilities other than j. This creates the possibility of "multiple-path" production (see Figure 11.5). For instance, in Figure 11.5-(b), production is started in period 1 at both facilities 1 and 2, while the former is used to satisfy demands in periods 2 and 3, and the latter is produced for period 1. This is possible when the production cost for facility 1 in period 1 is higher than that of the facility 2, but the combination of production and holding costs for periods 2 and 3 is lower in facility 1 than that of facility 2. Note that this particular case violates condition (i) and our shortest path algorithm will not identify this as a solution. Similarly, in Figure 11.5-(a),(c),(d), condition (ii) is violated since production takes place in one facility despite the fact that the other facility holds inventory.

Proposition 11.6 *When it is optimal to start a production at facility j in period t for a future period $t+l$ $(l > 0)$ while the demands in periods $t,\ldots,t+l-1$, are satisfied by facilities other than j (none of j's inventory is consumed during $t,\ldots,t+l-1$), then subproblem (P_i) can be no longer solved by our shortest path algorithm on the given network.*

Proof: It is sufficient to make an observation that under the above situation either condition (i) or (ii) in Proposition 11.5 will be violated. It is known from the Leontief structure that b^i_{jt} for a facility j can be either D^i_t or 0. Suppose the demand for period t is satisfied by facility k but not facility j, i.e., for facilities j and k, $b^i_{jt} = 0$ and $b^i_{kt} = D^i_t$, respectively. Since $x^i_{jt} + I^i_{jt-1} - I^i_{jt} = b^i_{jt}$, $\forall j, t$ it is possible to have (Case 1) $(x^i_{jt} = I^i_{jt} = D^i_{t+l}$ and $I^i_{jt-1} = 0)$ and $(x^i_{kt} = D^i_t$ and $I^i_{kt} = I^i_{kt-1})$

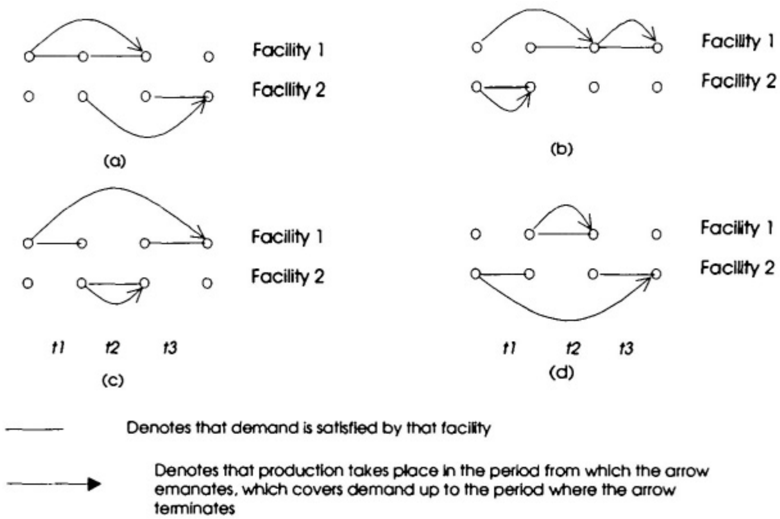


Figure 11.5. Four examples cases where the Shortest Path Algorithm may fail

(violation of condition (i)) (Case 2) ($x^i_{jt} = 0$ and $I^i_{jt} = I^i_{jt-1} = D^i_{t+l}$) and ($x^i_{kt} = D^i_t$ and $I^i_{kt} = I^i_{kt-1}$) (violation of condition (ii)) Recall that conditions (i) and (ii) in Proposition 11.5 are as follows:

(i) $x^i_{jt}x^i_{kt} = 0, \ \forall i, j, k \neq j, t,$

(ii) $x^i_{jt}I^i_{kt-1} = 0, \ \forall i, j, k, t.$

In *Case 1* above, it is clear that condition (i) is violated while in *Case 2,* condition (ii) is violated. $\diamondsuit$

The fact that the shortest path algorithm is not always optimal for subproblem (P_i) should not come as a surprise as Theorem 11.2 shows that the generalized cost version is *NP-complete.* Nevertheless, the shortest path algorithm is a valid heuristic for subproblem (P_i). In the computational experiments, we will show that the shortest path algorithm is an effective heuristic, and yields optimal or near-optimal solutions in most test cases.

3.2 Uncapacitated Single-Item, Multi-Facility Model with Transportation Costs

Recall that in the multicommodity flow network (Figure 11.2) and the single-item subnetwork (Figure 11.3) a *demand arc* $A_d \subset A$ goes from facility j in period t to demand point d^i_t that is characterized by

$(b^i_{jt}, r^i_{jt}, cap_{jt})$ where b^i_{jt} represents facility j's contribution to demand D^i_t, r^i_{jt} represents the per unit *transportation cost* from facility j to the demand point i, and cap_{jt} is the transportation capacity in period t. In Section 3.1 we assumed the transportation cost $r^i_{jt} = 0$ and $cap_{jt} = 4$. We will now consider the single-item subproblem where the transportation cost is positive but we will not consider transportation capacity. The subproblem can be stated as follows
(P_i')

$$Minimize \sum_{t=1}^{T} \sum_{j=1}^{J_i} (c_{jt}x_{jt} + K_j\delta_j + h_{jt}I_{jt} + r_{jt}b_{jt})$$

s.t.

<*mass balance constraints for commodity* i (11.2)>
<*setup constraint for commodity* i (11.4)>
<*nonnegativity and integrality constraints for commodity* i>

Theorem 11.7 *The uncapacitated, single-item, multiple-facility problem with non-negative transportation cost* r_{jt} *is NP-complete.*

Proof: See Appendix.

Algorithm Complexity

We now explore possible algorithms for problem (P_i') and their complexity as related to the number of facilities and number of periods. A first observation is made by viewing the single-item, multi-facility problem depicted in Figure 11.3 as a Directed Steiner Problem (Leibling, 1999): the subnetwork represents the problem graph, while the source, the root of each production subnet, and the demand nodes (on the right) are "compulsory" nodes, and the remaining are "optional" or Steiner points. The Directed Steiner Tree Problem is to find an arborescence of minimal total length that spans all compulsory nodes using one or more Steiner points. If arc lengths can be determined *a priori* using production, inventory, and transportation costs, an optimal Steiner solution corresponds to an optimal solution to (P_i'). It is known that a dynamic programming algorithm (Dreyfus and Wagner, 1972) solves the Steiner Problem in $O[(N_c + N_s) \cdot 3^{N_s} + (N_c + N_s)^2 \cdot 2^{N_s}]$ where N_s is the number of Steiner points (i.e., number of facilities x number of periods) and N_c is the number of compulsory points (i.e., number of periods). While this observation suggests that a dynamic programming algorithm constructed in this fashion is exponential in the number of facilities *and* periods, using the well-known Wagner-Whitin results, we could improve the complexity significantly. This is stated as follows.

Theorem 11.8 *There exists an algorithm for the uncapacitated, single-item, multiple-facility problem with non-negative transportation cost (Problem $(P_i'))$ that is polynomial in the number facilities and exponential in the number of periods.*

Proof: Consider a simple algorithm as follows for problem (P_i'):

1 Make all combinations of assignments of $D_1, \ldots, D_T$ to facilities $j = 1, \ldots, m$ (this takes $O(m^T)$)

2 Given each assignment, solve for each facility $j = 1, \ldots, m$ an uncapacitated, single-item, single-facility lot-sizing problem using the Wagner-Whitin Algorithm (this takes $m \cdot O(T \log T)$)

The algorithm has complexity $O(m^{T+1} \cdot T \log T)$, which is polynomial in the number of facilities but exponential in the number of periods. ◇

3.3 Special Cost Structure

Despite these negative results, the shortest path algorithm remains a viable option under certain cost structures. Specifically, when transportation and production costs are fixed, when there is no incentive to hold inventory for more than four periods without starting a new setup, and when holding inventory in a facility j for two periods always costs more than holding inventory elsewhere for one period (i.e., when holding costs are not dramatically different), all cases stated in Proposition 11.6 and Figure 11.5 disappear.

Proposition 11.9 *The uncapacitated, single-item, multiple-facility problem with non-negative transportation cost (P_i') can be solved in polynomial time using a shortest path algorithm if the following conditions hold:*

(i) *The transportation and the production costs are fixed across facilities and periods*

(ii) *There is no setup that would last more than four periods, i.e.,* $\sum_{\tau=t}^{t+3} h_{j\tau} D_\tau \geq K_j, \forall j$

(iii) *Inventory holding costs of any two facilities j and k satisfy the relationship $h_{jt} + h_{j,t+1} \geq h_{k,t+1}$ $\forall j, k, t$.*

It should be intuitive when conditions (i)-(iii) above hold in the data set, the condition in Proposition 11.6 is eliminated, while conditions (i) and (ii) in Proposition 11.5 will be satisfied, i.e., the shortest path algorithm will produce optimal solution. These conditions can be thus used as a test to be performed on the data set before calculations start. It is quite likely that even when the above conditions are violated, the shortest path algorithm remains an effective heuristic. To put this point under solid empirical testing, we conducted two sets of computational experiments: the first experiment simply compares the optimal solutions for (P_i) obtained by an MIP (Mixed Integer Programming) solver versus the solution provided by a shortest path algorithm. The results are summarized in Section 3.4. The *second* set of experiments is more complicated; here we test the quality of the single-item shortest path algorithm as a subproblem heuristic for the multi-item problem (P). This will be detailed in Section 4.

3.4 Performance of the Shortest Path Algorithm

To instantiate the insights gained from the analytic results, we conduct intensive empirical testing by varying setup, holding, production, and transportation costs, the number of facility and period combinations, and levels of demand lumpiness. This results in 17,500 instances. We first generate a nominal case as follows:

Setup cost, K_j ~ *Uniform[1500,3500]*

Production cost, $c_{jt} \sim$ [illegible]

Transportation cost, r_{jt} ~ *Uniform [5,15]*

Holding cost, $h_{jt} \sim$ [illegible]

Number of facilities: *4*

Number of periods: *6*

Demand, $D_t \sim$ [illegible]

No demand lumpiness

Demand lumpiness is used to generate clustered demand patterns to more closely resemble practical situations. While generating random demand for D_t in period t, we first generate a random number between 0 and 1; if it is less than the current lumpiness threshold ζ, then D_t is set to 0, otherwise, D_t is set to the generated demand. We adjust the demand levels such that the expected demand remains the same as the base case.

By varying one or more factors from the nominal case, we generate three groups of test problems as follows:

(Group One): Fixed setup costs across all periods, i.e., the setup cost is generated once and fixed for all periods (by varying the following cost factors, we generate 65 sub-groups where 100 replications are generated for each subgroup)

a1-a10: varying the *setup cost* levels from *Uniform [0,500]* to *Uniform [5000,10000]*

a11-a20: varying the holding cost levels from *Uniform [1,5]* to *Uniform[100,200]*

a21-a30: varying the production cost levels from *Uniform[l,5]* to *Uniform [100,200]*

a31-a40: varying the transportation cost levels from *Uniformf[1,5]* to *Uniform[100,200]*

a4 l-a60: varying the number of facilities (J_i)/number of periods (T) combinations, or $| J_i | \times | T |$, where $| J_i | = 2, 3, 4, 5$, and 6, and $| T | = 6, 8, 10$, and 12

a61-a65: varying the demand lumpiness threshold $\zeta = 0, 0.2, 0.3, 0.4, 0.5$

(Group Two): Variable setup cost across periods, i.e., setup costs are randomly generated for each period t. Again, 65 sub-groups are generated as above with 100 replications each.

(Group Three): Fixed production and transportation costs, i.e., the production and transportation costs are fixed to their expected value (10). This eliminates subgroups *a21-a40* above, resulting in 45 subgroups with 100 replications each.

A summary of the test results is given in Figure 11.6. As shown in the figure, except for subgroups *a37-a40,* the shortest path algorithm produces nearly identical solutions to the MIP (i.e., less than 3 occurrences per 100 instances, with a maximum solution deviation less than 0.2%). When more significant deviations are observed in subgroups *a37-a40* (cases where the transportation costs are high at *Uniform [40,80],* ..., *Uniform [100,200],* respectively), although the number of occurrences is much more pronounced (up to 49 per 100 instances), the maximum amount of deviation is still under 0.5% in all cases. Also observed from the experiments is that more significant deviation results (between MIP and shortest path) when the setup costs are variable across periods *(Group Two).* Now consider the *Group Three* results: there is no deviation between shortest path and MIP solutions in any of the 4,500 cases tested. (Of course this is only an empirical observation.) Consider this result along with Proposition 11.9. It is interesting to note that when the production and transportation costs are fixed (condition (i) in the proposition), even when conditions (ii) and (iii) do not hold, the shortest path algorithm produced optimal solutions. These results confirm our insight from Proposition 11.6 that cases exist where the shortest path algorithm produces sub-optimal solutions; upon examining these cases, we found that they are indeed caused by circumstances as depicted in Figure 11.5. On the other hand, the empirical results show that the shortest path algorithm is a very effective heuristic for the single-item, multi-facility subproblem. In the following section, we consider this single-item subproblem in the context of the multi-item problem.

4. A Solution Methodology for the Multi-Item, Multi-Facility Problem (P)

We propose a Lagrangean Decomposition scheme for solving the multi-item, multi-facility problem (P) where the single-item problem (P_i) is a subproblem. Lagrangean Decomposition, as described in Guignard and Kim (1987), has been applied to a variety of NP-hard problems including multi-item single-facility lot-sizing problems (Thizy 1991). A main advantage of Lagrangean Decomposition over the better known Lagrangean Relaxation is that the theoretical Lower Bound obtained from Lagrangean Decomposition is at least as tight as that from Lagrangean Relaxation. We start our exposition by first listing the mass balance constraints 11.2 explicitly:

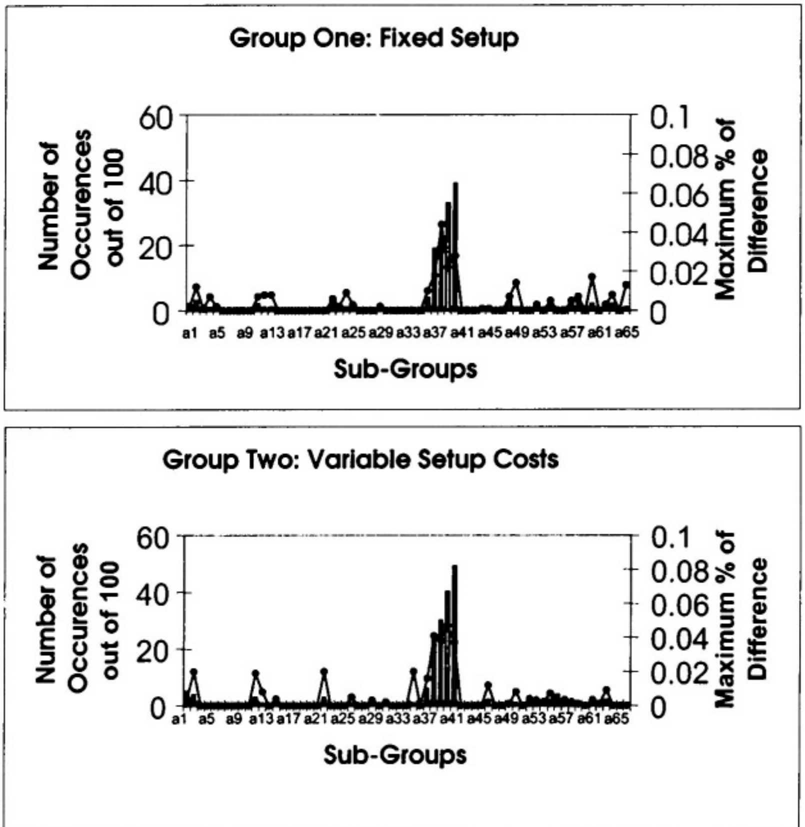
$$\sum_{t=1}^{T} D_t^i = \sum_{j=1}^{J_i} f_j^i, \qquad \forall i \in N, \tag{11.6}$$

$$f_j^i = \sum_{t=1}^{T} x_{jt}^i, \qquad \forall i \in N, j \in J_i, \tag{11.7}$$

$$x_{jt}^i + I_{jt-1}^i = b_{jt}^i + I_{jt}^i, \quad \forall i \in N, (j,t), \in A \tag{11.8}$$

$$\sum_{j=1}^{J_i} b_{jt}^i = D_t^i, \qquad \forall i \in N, t \in T. \tag{11.9}$$

If we assume that the system must return to its initial inventory at the end of the planning horizon, i.e., $I_0 = I_T = 0$, we may simplify the above mass balance constraints. From (11.8) and (11.9) we have,



Group Three: Fixed Production and Transportation Costs

In all 4,500 instances in this group, the Shortest Path Algorithm produces the same solution as the Mixed Integer Program, i.e., the number of occurrences and the maximum % of difference are 0 in all cases.

Figure 11.6. Comparing the Shortest Path and the MIP Solutions (the bars represent the number of different solutions out of 100, while the line graph represents the maximum % of differences)

$$D_t^i = \sum_{j=1}^{J_i} (x_{jt}^i + I_{jt-1}^i - I_{jt}^i), \ \forall i, t, \ thus$$

$$\sum_{t=1}^{T} D_t^i = \sum_{t=1}^{T} \sum_{j=1}^{J_i} (x_{jt}^i + I_{jt-1}^i - I_{jt}^i) \tag{11.10}$$

but since $I_0 = I_T$

$$\sum_{t=1}^{T} I_{jt-1}^i = I_0 + \sum_{t=1}^{T-1} I_{jt}^i = \sum_{t=1}^{T-1} I_{jt}^i + I_T = \sum_{t=1}^{T} I_{jt}^i \tag{11.11}$$

so we can rewrite (11.10) as

$$\sum_{t=1}^{T} D_t^i = \sum_{j=1}^{J_i} \sum_{t=1}^{T} x_{jt}^i, \forall i \in N \tag{11.12}$$

Note that (11.12) implies constraints (11.6) and (11.7). This allows us to consider problem (P) with only two sets of balance constraints (11.8) and (11.9) since (11.6) and (11.7) will be satisfied automatically. Now, consider the objective function of (P):

$$Minimize\ z = \sum_{t=1}^{T} \sum_{i=1}^{n} \sum_{j=1}^{J_i} (c_{jt}^i x_{jt}^i + K_j^i \delta_{jt}^i + h_{jt}^i I_{jt}^i + r_{jt}^i b_{jt}^i)$$

Since $b_{jt}^i = x_{jt}^i + I_{jt-1}^i - I_{jt}^i$ (from (11.8)), the objective can be rewritten as follows:

$$Minimize\ z = \sum_{t=1}^{T} \sum_{i=1}^{n} \sum_{j=1}^{J_i} ((c_{jt}^i + r_{jt}^i) x_{jt}^i + (h_{jt}^i - r_{jt}^i + r_{jt+1}^i) I_{jt}^i + K_j^i \delta_{jt}^i)$$

If we set $\hat{c}_{jt}^i \equiv c_{jt}^i + r_{jt}^i$ and $\hat{h}_{jt}^i \equiv h_{jt}^i - r_{jt}^i + r_{jt+1}^i$ to be the generalized production and holding costs, the objective becomes:

$$Minimize\ z = \sum_{t=1}^{T} \sum_{i=1}^{n} \sum_{j=1}^{J_i} ((\hat{c}_{jt}^i x_{jt}^i + \hat{h}_{jt}^i I_{jt}^i + K_j^i \delta_{jt}^i)$$

The basic idea of our decomposition is to separate the multicommodity flow problem (P) into two subproblems: the first with the capacity and the mass balance constraints, but not the setup constraints; the second is a commodity-decomposable subproblem with the mass balance and the setup constraints. The latter defines separable single-item problems (P_i), which have special structure as analyzed in Section 3.

In this decomposition, the first subproblem is a linear program and the second is a collection of MIP problems. If we make use of the shortest path algorithm for each MIP (as a heuristic), all subproblems are easy to solve. To demonstrate this solution methodology we use a slightly simplified formulation of problem (P) by dropping the transportation term in the objective (as above), and assuming that a setup does not consume capacity (dropping (11.3)). We then restate the multi-facility production problem (P) with duplicated variables as follows:
(P')

$$Minimize\ z = \sum_{t=1}^{T}\sum_{i=1}^{n}\sum_{j=1}^{J_i}((\hat{c}^i_{jt}x^i_{jt} + \hat{h}^i_{jt}I^i_{jt} + K^i_j\delta^i_{jt}) \qquad (11.13)$$

s.t.

$$\sum_{i=1}^{n}\beta^i_{jt}x^i_{jt} \le u_{jt}, \qquad \forall (j,t) \in A, \qquad (11.14)$$

$$x^i_{jt} + I^i_{jt-1} - I^i_{jt} \ge 0, \qquad \forall i \in N, (j,t) \in A, \qquad (11.15)$$

$$\sum_{j=1}^{J_i}(x^i_{jt} + I^i_{jt-1} - I^i_{jt}) = D^i_t, \qquad \forall i \in N, t \in T, \qquad (11.16)$$

$$x^i_{jt} = xx^i_{jt}, \qquad \forall i \in N, j \in J, t \in T, \qquad (11.17)$$

$$I^i_{jt} = II^i_{jt}, \qquad \forall i \in N, j \in J, t \in T, \qquad (11.18)$$

$$xx^i_{jt} + II^i_{jt-1} - II^i_{jt} \ge 0, \qquad \forall i \in N, (j,t) \in A, \qquad (11.19)$$

$$\sum_{j=1}^{J_i}(xx^i_{jt} + II^i_{jt-1} - II^i_{jt}) = D^i_t, \qquad \forall i \in N, t \in T, \qquad (11.20)$$

$$xx^i_{jt} \le M\delta^i_{jt}, \qquad \forall i \in N, j \in J, t \in T, \qquad (11.21)$$

$$x^i_{jt}, I^i_{jt} \ge 0, \qquad \forall i \in N, j \in J, t \in T, \qquad (11.22)$$

$$xx^i_{jt}, II^i_{jt} \ge 0, \qquad \forall i \in N, j \in J, t \in T, \qquad (11.23)$$

$$\delta^i_{jt} \in (0,1), \qquad \forall i \in N, j \in J, t \in T. \qquad (11.24)$$

In the above formulation, we make copies of the variables x^i_{jt} and I^i_{jt} as xx^i_{jt} and II^i_{jt}. We then use the copies to split the original constraints into two sets of constraints: (11.14)-(11.16), (11.22) and (11.19), (11.20), (11.21), (11.23), (11.24) plus the linking constraints (11.17)-(11.18). It should be clear that $(\mathrm{P}) \equiv (P')$. We then separate *(P')* by relaxing the linking constraints and placing them in the objective function with Lagrangean multipliers λ^x_{ijt} and λ^I_{ijt}. This yields the following subproblems:

Resource Subproblem:

$$Minimize\ z_1 = \sum_{t=1}^{T}\sum_{i=1}^{n}\sum_{j=1}^{J_i}((\hat{c}^i_{jt} + \lambda^x_{ijt})x^i_{jt} + (\hat{h}^i_{jt} + \lambda^I_{ijt})I^i_{jt}) \equiv z_1(\lambda)$$

s.t. (11.14), (11.15), (11.16), and (11.22).

***n* Product Subproblems:**

$$Minimize\ z_2 = \sum_{i=1}^{n}(\sum_{t=1}^{T}\sum_{j=1}^{J_i}(-\lambda^x_{ijt}xx^i_{jt} - \lambda^I_{ijt}II^i_{jt} + K^i_j q^i_{jt})) \equiv \sum_{i=1}^{n} z^i_2(\lambda)$$

s.t. (11.19), (11.20), (11.21), (11.23), and (11.24).

Note that under the general framework of Lagrangean Decomposition, constraints (11.15) and (11.16) do not need to be duplicated for *both* subproblems, i.e., these constraints can be assigned to *either* subproblem. However, our computational experience indicates that as long as the added constraints do not add computational burden to the subproblems, constraint duplication improves the speed of convergence and yields better lower bounds since the solutions proposed by the subproblems tend to be similar. A lower bound to problem *(P')* given the Lagrangean multiplier set λ is as follows:

$$LB_\lambda(P') = v(z_1(\lambda)) + \sum_{i=1}^{n} v(z^i_2(\lambda)),$$

where $v(.)$ denotes the optimal value of the objective. Note that the resource subproblem is a linear program, and the product subproblem z^i_2 has similar structure to problem (P_i) described earlier. Prom the lower bound solution, an upper bound for problem *(P')* can be generated using the following feasibility restoration routine: given the solution for the resource subproblem we add setups for the periods where production is nonzero, i.e., we set q^i_{jt} to 1 whenever x^i_{jt}. We then calculate the objective function using the original cost function. This results in an upper bound for the original problem. The lower bound can be maximized by searching for the set of Lagrangean multipliers λ that maximize the Lagrangean dual. Both dual ascent and subgradient search methods can be used for this task. In this paper, we use the latter approach, which is summarized in Section 4.2. As we will demonstrate in the computational section, we can achieve solutions with very small duality gaps using the bounds and the search algorithm.

4.1 Managerial Insights Related to the Decomposition

Our choice of Lagrangean Decomposition is not purely motivated by computing. The decomposition of the multi-facility production model into a resource subproblem and multiple product subproblems has interesting managerial implications. As recognized by several researchers (c.f., Jörnsten and Leisten 1994, Burton and Obel 1984), mathematical decomposition often leads to insights for general modeling strategies or even new decision structures. The decomposition suggested earlier allows further analysis concerning modeling flexibility in the context of multi-facility manufacturing planning. Suppose we consider each product subproblem as a decision problem for a *product* manager and the resource subproblem as a decision problem for a *production* manager overseeing multiple facilities. Thus, the decomposition can be viewed as a decision system where *product* managers, each responsible for a product, compete for resource capacity available from manufacturing facilities. The *production* manager, on the other hand, represents the interests of efficiently allocating resources from multi-facilities to the products. Clearly the solutions proposed by the production manager (x, I) do not agree with the collective solution proposed by the product managers (xx, II). A search based on Lagrangean multipliers essentially penalizes their differences, while adjusting the penalty vector iteratively. This process stops when the degree of disagreement (the duality gap) is acceptably low, or when further improvement is unlikely.

The above viewpoint is useful in evaluating the flexibility model (P) represents. First, it should be clear that each product subproblem (P_i) could be customized to represent the distinctive needs of each product. So long as its basic network structure is maintained there will be no additional computational burden. Similarly, as long as the resource subproblem remains a linear program, it can be customized with various facility submodels each reflecting the distinct production structure of a facility. However, a different constraint duplication strategy may be necessary when changes are made to the base model.

4.2 The Subgradient Search Algorithm

In this section, we summarize the subgradient search algorithm used to adjust the Lagrangean multipliers. At each iteration s, we calculate Lagrangean multipliers using the following equations:

$$\lambda_{ijt}^{x,s+1} = \lambda_{ijt}^{x,s} + u^s(x_{jt}^{i,s} - xx_{jt}^{i,s}),$$
$$\lambda_{ijt}^{I,s+1} = \lambda_{ijt}^{I,s} + u^s(I_{jt}^{i,s} - II_{jt}^{i,s}), \qquad (11.25)$$

where

$$u^s = \frac{\gamma_s(UB_s^* - (v(z_1(\lambda^s)) + \sum_{i=1}^{n} v(z_2^i(\lambda^s)))}{\sum_{t=1}^{T}\sum_{i=1}^{n}\sum_{j=1}^{J_i}((x_{jt}^{i,s} - xx_{jt}^{i,s})^2 + (I_{jt}^{i,s} - II_{jt}^{i,s})^2)}, \tag{11.26}$$

and

$\gamma_s =$ a scalar set to 1 and reduced by half if the lower bound fails to improve after a fixed number of iterations

$UB_s^* =$ the best upper bound obtained up to iteration s

In our testing, we terminate the algorithm after a prespecified number of iterations. The best upper bound obtained at the end of the iterations provides the heuristic solution to the problem. We summarize the algorithmic steps as follows:

Step 1: *Initialize* s, λ, u, γ *and UB*.*

Step 2: *Solve the resource and the product subproblems. Compute the lower bound* $(v(z_1(\lambda^s)) + \sum_{i=1}^{n} v(z_2^i(\lambda^s))$ *for the current iteration,* s.

Step 3: *Compute an upper bound* UB_s *from the optimal solution of the current resource subproblem Min* $z_1(\lambda_s)$. *If* $UB_s < UB_{s-1}^*$, *set* $UB_s^* \leftarrow UB_s$.

Step 4: *Update the multipliers using equations (11.25) and (11.26)*

Step 5: *Stop if a prespecified iteration limit is reached. Otherwise go to Step 2.*

5. Computational Testing

5.1 Effectiveness of Lagrangean Decomposition and Subgradient Search

We implemented the subgradient search algorithm using the mathematical programming language AMPL with the CPLEX solver. The experiments are conducted on a Pentium-200 personal computer with 64Meg RAM. To test the effectiveness of the Lagrangean decomposition and the subgradient search algorithm, we generate 450 test instances with 30 distinctly different problem characteristics. The product subproblems are solved using the CPLEX MIP solver. For each test instance we use 150 iterations of subgradient search and record the best lower and upper bounds at the end of the iteration to compute the duality gap. We first generate two sets of *nominal case problems* (*C 1*) with 4 facilities and 6 periods. The first set assumes fixed transportation (FT) costs while the second set assumes variable transportation (VT) costs. The demand and cost parameters used for the nominal case are summarized in the footnotes of Table 11.1. As shown, the production, inventory and setup costs as well as item demands are randomly generated using a Uniform distribution. To generate capacity we use the following procedure: we first calculate cumulative demands by adding the randomly generated demands of all items up to period t, for $t = 1, \ldots, T$. For the first period, we multiply the total demand for the period by some constant (≥ 1). We then use this number as the total capacity available in the period and assign a fraction to each facility. For the coming periods, total capacities assigned for the previous time periods are subtracted from the cumulative demand of that period and then multiplied by some constant to generate the capacity. Using this procedure, we may generate relatively challenging (but feasible) test problems with tight capacity constraints. For the nominal test problems these constants are set at 1.3. Fifteen replications are assigned to each case. We then alter the nominal cases by changing problem characteristics and sizes to generate 14 additional cases (*C2-C15*) each including the fixed and variable transportation cost (FT and VT) cases and each are repeated for 15 replications. This results in 450 test instances and the average duality gaps are summarized in Table 11.1. For simplicity, we assume β^i_{jt} (the consumption rate of facility j's resource by item i at period t) is equal to 1. We also assume the starting and ending inventory to be zero. For most of the test problems the lower bound increases significantly in the first 20 iterations whereas the upper bound improves slowly. There appears to be a strong correlation between the quality of the lower and the upper bounds, i.e., when the lower bound obtained is tight, the upper bound restored from the lower bound solution is also of higher quality. We observed a quite consistent convergence pattern throughout all test problems. Convergence typically occurs quite early resulting in a very small duality gap. We summarize the observations from Table 11.2 as follows:

1 As shown in the table, setup cost appears to have a significant effect on the duality gap. Low setup instances have an average gap of 1.72% and 1.39% compared to 7.25% and 6.51% for the high setup instances. This result is not surprising since increased setup costs widen the gap between the resource subproblem (which is an LP ignoring the setup cost) and the product subproblems. On the other hand, since the original problem is a mixed integer

%	C1	C2	C3	C4	C5	C6	C7	
FT	4.07	1.72	7.25	1.82	0.91	4.36	2.27	
VT	4.02	1.39	6.51	1.87	0.94	4.00	2.54	
	C8	C9	C10	C11	C12	C13	C14	C15
FT	3.73	4.93	0.84	2.09	3.10	12.3	7.05	25.9
VT	2.78	5.13	0.84	2.01	2.64	12.4	6.83	25.9

Duality gaps are calculated as *(UB–LB)/UB* *100% each table entry is averaged over 15 replications.

C1- Nominal Case: Demand~U(0,200), Setup cost~U(1500,3500), Holding cost~U(5,15), Production cost~U(5,15), 30 items, 4 facilities, 6 periods, Capacity tightness factor =1.3, Transportation cost:10 for FT, U~(5,15) for VT.

The following cases represent variations from the nominal cases by the indicated factor(s):

C2- Low setup where setup cost ~Uniform (0,1000).

C3- High Setup where setup cost~Uniform (4000,8000).

C4- High Production Costs where Production cost~Uniform (40,80).

C5- Very High Production Cost where Production cost~Uniform(100,200).

C6- Lumpy demand: expected demand is 100 but there is a 0.3 probability that demand is 0.

C7- Low demand variability where Demand ~Uniform (50,150).

C8- Loose capacity: capacity tightness factor is set at 1.8.

C9- Tight capacity: capacity tightness factor is set at 1.15.

C10- Number of facilities=1.

C11- Number of facilities=2.

C12- Number of facilities=3.

C13- Number of items=10.

C14- Number of items=20.

C15- Projected worst case: Lumpy demand, High setup cost, Tight capacity, number of items=10.

Table 11.1. Duality gaps for 450 test instances over various cost structures and problem sizes

program with binary setup variables, as the setup costs increase the problem behaves closer to a combinatorial problem then an LP.

2 The number of facilities appears to have an effect on the duality gap as well. Consider cases *C1* and *C10-C12.* As the number of facilities increases we observe a monotonic increase in duality gap as well. This result is useful in that making alternative facility production decisions is a unique feature of our model. The results suggest that the added dimension has a noticeable effect on the difficulty of the problem. On the other hand, it also shows that the proposed algorithm is quite effective in solving the traditional single-facility problems (*C10*).

3 The effect of capacity levels is much less pronounced. This may be due to the fact that the capacity generation procedure produces relatively tight capacity in all instances. Since the difference between non-capacitated and capacitated lot sizing models is well known, we did not make an attempt to further loosen the capacity.

4 Increasing the number of items appears to have an effect on the duality gap as well. Problems with a larger number of items appear to have a smaller duality gap.

5.2 Effectiveness of the Shortest Path Algorithm as a Subproblem Heuristic

The results in Table 11.1 are produced by Lagrangean Decomposition using subgradient search where each single-item, multi-facility subproblem is solved as an MIP. As demonstrated in Section 3.4, the shortest path algorithm can be an effective heuristic for the single-item subproblem (the solution never deviates from the optimal by more than 0.5%). We are interested in the effectiveness of the shortest path algorithm as a heuristic in the context of subgradient search. The potential savings in computing time is significant as the subgradient search algorithm must solve $|N|$ single-item subproblem at each iteration; this results in up to 3,000 calls to the subproblem (i.e., 20 items, and 150 iterations). However, it should be recognized that the subgradient search method might not work properly when the shortest path algorithm is used to solve the subproblems. This is the case when the shortest path algorithm produces suboptimal solutions to the (capacity-relaxed) subproblem *and* the solution values (of the relaxed subproblem) exceed the optimum of the original (capacitated) problem. In Table 11.2, we compare the bounds generated for the multi-item multi-facility problem (P') when

	C1	C2	C3	C4	C5	C6	C7
# of times* UB(SP)<UB(MIP)	14	9	13	12	13	10	15
% difference in UB values**	0.40	0.43	1.02	0.25	0.24	0.97	0.49
# of times* UB(SP)>UB(MIP)	1	6	2	3	2	5	0
% difference in UB values**	2.65	0.18	0.03	0.45	0.04	0.25	0.00
# of times* UB(SP)=UB(MIP)	1	0	0	0	0	0	0
# of times* LB(SP)>LB(MIP)	7	4	10	4	4	5	4
% difference in LB values**	0.03	0.18	0.04	0.01	0.01	0.04	0.03
# of times* LB(SP)<LB(MIP)	5	7	4	3	6	6	3
% difference in LB values**	0.02	0.02	0.03	0.02	0.01	0.02	0.01
# of times* LB(SP)=LB(MIP)	3	4	1	8	5	4	8

	C8	C9	C10	C11	C12	C13	C14	C15
# of times* UB(SP)<UB(MIP)	14	8	11	13	12	14	11	12
% difference in UB values**	0.51	0.57	0.20	0.00	0.00	1.28	1.04	1.96
# of times* UB(SP)>UB(MIP)	1	7	4	2	3	1	4	3
% difference in UB values**	0.45	0.64	0.14	0.03	0.34	0.04	0.59	0.61
# of times* UB(SP)=UB(MIP)	0	0	0	13	12	0	0	0
# of times* LB(SP)>LB(MIP)	6	3	0	0	1	7	7	5
% difference in LB values**	0.01	0.06	0.00	0.00	0.01	0.03	0.08	0.06
# of times* LB(SP)<LB(MIP)	5	11	0	0	0	7	4	9
% difference in LB values**	0.02	0.04	0.00	0.00	0.00	0.06	0.03	0.23
# of times* LB(SP)=LB(MIP)	4	1	15	15	14	1	4	1

* the number of occurrences out of 15 instances

** averaged over the number of occurrences in the cells above

cases C1-C15 are equivalent to cases defined in Table 1

Table 11.2. Comparing the quality of bounds for the multi-item problem when the single-item problems are solved by the MIP and the shortest path algorithm

MIP and the shortest path algorithm are used to solve the single-item subproblems.

It should be evident that when the shortest path algorithm is used to solve the single-item subproblem, it produces nearly identical lower and upper bounds for the multi-item problem. The percentage difference in lower bounds produced by shortest path and MIP never exceeds 1% in all but 2 cases (out of 450 instances) and, in most cases, the difference is less than 0.02%. Even in the worst cases (case *C15*) where the

	C16	C17	C18	C19	C20
# of times UB(SP)<UB(MIP)*	12	0	2	3	1
*% difference in UB values***	0.29	0.00	0.03	0.04	0.01
# of times UB(SP)>UB(MIP)*	3	15	10	10	14
*% difference in UB values***	0.16	0.24	0.08	0.05	0.09
# of times UB(SP)=UB(MIP)*	0	0	3	2	0
# of times LB(SP)>LB(MIP)*	0	0	0	2	0
*% difference in LB values***	0.00	0.00	0.00	0.03	0.00
# of times LB(SP)<LB(MIP)*	15	15	15	12	15
*% difference in LB values***	0.10	0.20	0.06	0.06	0.12
# of times LB(SP)=LB(MIP)*	0	0	0	1	0

* the number of occurrences out of 15 instances

** averaged over the number of occurrences in the cells above

C16: High Transportation Costs ~Uniform(40,80)

C17: Very High Transportation Costs ~Uniform(100,200)

C18: Low Setup ~Uniform (0,1000), High Transportation Costs ~Uniform(40,80)

C19: Low Setup ~Uniform (0,1000), High Transportation Costs ~Uniform(40,80), Lumpy Demand

C20: Low Setup ~Uniform (0,1000), Very High Transportation Costs ~Uniform(100,200)

Table 11.3. Comparing the quality of bounds for specially generated worst cases

shortest path is expected to generate different solutions than MIP, the actual difference averaged at 0.23%. To further test these worst cases, we generate an additional five classes of test problems (*C16-C20*) by intentionally increasing the transportation costs, lowering the setup costs, and employing combinations of the two. These additional results are given in Table 11.3. As can be seen from the table, the difference in bound quality between shortest path and MIP is still less than 0.29% on average. However, MIP does generate better upper bounds and tighter lower bounds in these cases.

Another point of interest is how often during the subgradient search the shortest path solution deviates from the solution generated by MIP. The statistics collected on this particular measure are given in Figure 11.7.

A main incentive for studying the single-item, multi-facility shortest path algorithm is the potential computational saving when solving the multi-item problem. We conduct an additional set of experiments that

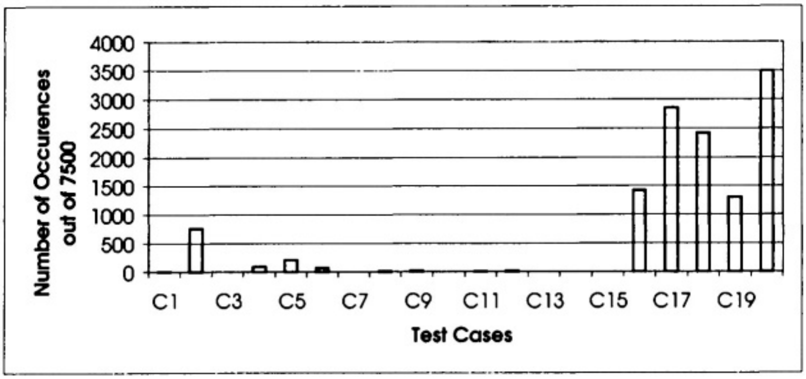


Figure 11.7. The average frequency when the shortest path solution deviates from MIP during the subgradient search (150 iterations × 50 Items = 7500 instances under each test case, averaged over 15 replications)

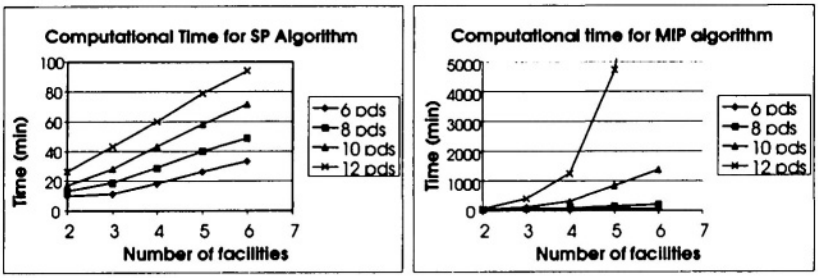


Figure 11.8. Comparison Between Shortest Path Algorithm and MIP in the Multi-item problem (each data point represent 150 iterations of subgradient search)

compare the computational efficiency of the shortest path algorithm with MIP. While it is obvious that the shortest path algorithm will take much less time in each single-item problem, we are interested in knowing the real savings in the context of the multi-item, multi-facility problem, and the effect of increasing the number of facilities and number of periods. We generate 20 multi-item problems with 2, 3, 4, 5, or 6 facilities, and 6, 8, 10 or 12 periods. Each problem is solved using subgradient search with 150 iterations implemented in AMPL/CPLEX. The total computer time is recorded in CPU minutes, and the results are summarized in Figure 11.8. As shown in the figure, when the shortest path algorithm is used for the single item subproblem, the computer time increases in a near linear fashion as the number of facilities increases. On the other hand, when we solve an MIP for each subproblem, the computer time increases much more dramatically. An exponential growth in computer time can be observed when testing the 12-period problem, where we can only solve up to 5 facilities with the 5,000 CPU-minute limit. For problems that are solvable using the MIP subproblem we observe a much sharper increase in CPU time as the number of periods increases. This is consistent with the complexity insights given in Theorem 4.

Acknowledgments

This research is supported, in part, by National Science Foundation grant DMI-9634808 and a grant from Agere Systems (formerly Lucent Technologies). We appreciate insightful comments from Professors Thomas Liebling and George Nemhauser on an earlier version of the formulation.

Appendix

Proof of Theorem 11.1. It is easy to verify that problem (P_i) satisfies the Leontief structure. The setup cost K_j^i can be incorporated into the production cost C_{jt}^i as a fixed charge function as below:

$$C_{jt}^i = \begin{cases} c_{jt}^i x_{jt}^i + K_j^i, & \text{if } x_{jt}^i > 0 \\ 0, & \text{if } x_{jt}^i = 0 \end{cases}.$$

Other constraints are linear while the objective function is concave. Thus model (P_i) has the following features: all nonnegative variables x, I, b, appear exactly once with a positive (+1) coefficient; in all other occurrences they have a negative (-1) coefficient. It follows that if more than one variable appears with a positive coefficient in the same constraints, then only one of these variables can be positive in an optimal solution, which results in the following conditions:

$$b_{jt}^i b_{kt}^i = 0 \; for \; t = 1, \ldots, T, j = 1, \ldots, J_i, k = 1, \ldots, J_i,$$

This condition states the *non-splitting property.* Similarly, the Leontief structure states the following:

$x^i_{jt} I_{jt-1} = 0$, $\forall i, j, t$ or no production of item i will take place when there is inventory in the same facility. ◇

Proof of Theorem 11.2. We first state the uncapacitated single-item production-planning problem over alternative facilities by writing out the mass balance constraints. We concentrate on the problem where inventory holding costs h are not restricted in sign, we call this problem *UCAP-N.*

(UCAP-N)

$$Minimize \sum_{t=1}^{T} \sum_{j=1}^{J_i} (c_{jt} x_{jt} + K_j \delta_{jt} + h_{jt} I_{jt})$$

s.t.

$$\begin{aligned}
x_{jt} + I_{jt-1} - I_{jt} = b_{jt}, \quad & \forall j \in J_i, t = 1, \ldots, T, \\
\sum_{j=1}^{J_i} b_{jt} = D_t, \quad & \forall t = 1, \ldots, T, \\
x_{jt} \le M \delta_{jt}, \quad & \forall j \in J_i, t = 1, \ldots, T, \\
x_{jt}, I_{jt} \ge 0, \quad & \forall j \in J_i, t = 1, \ldots, T, \\
\delta_t \in \{0, 1\} \quad & \forall t = 1, \ldots, T.
\end{aligned}$$

To show that *UCAP-N* is *NP-complete* we will show that the uncapacitated facility location (*UFL*) problem can be reduced to *UCAP-N.* Specifically, we will show that for any given instance of the *NP-complete* problem *UFL,* there is a corresponding instance of *UCAP-N.* The decision version of *UFL* is given as follows:

(Uncapacitated Facility Location) *Construct m facilities, and set $J_i = 1, \ldots, m$ for all demand points $i = 1, \ldots, n$. Each demand point can be assigned to one of the facilities (some facilities may not be possible for a specific demand point). When a demand point i is assigned to facility j (when y_{ji} is set to 1), a demand assignment cost d_{ji} is incurred. It is necessary that all demand points be assigned to some facility. When one or more demand points are assigned to a facility j there is a cost, f_j to open the facility; otherwise there is no cost. Given the above conditions, is there an assignment of demand points to facilities such that the total cost is lower than κ?*

For convenience, we also state the optimization version of the *UFL* problem as follows:
(UFL)

$$Minimize\ z = \sum_{i=1}^{n} \sum_{j=1}^{m} d_{ji} y_{ji} + \sum_{j=1}^{m} f_j \chi_j$$

s.t.

$$\begin{aligned}
\sum_{j=1}^{m} y_{ji} = 1, \quad & \forall i, \\
\sum_{i=1}^{n} y_{ji} \le \chi_j, \quad & \forall j, \\
y_{ji}, \chi_j \in \{0, 1\} \quad & \forall i, j.
\end{aligned}$$

For a given instance of the *UFL* problem we construct the following *UCAP-N instance:* Construct $T = n + 1$ periods including an initial period s before period 1 so that each demand point $1, \ldots, n$ corresponds to a planning period in *UCAP-N*. Construct $J_i = 1, \ldots, m \; \forall i = 1, \ldots, n$ facilities such that a facility in *UFL* correspond to a facility in *UCAP-N*. Define set up cost K_j for facility j according to facility opening cost f_j. Set $c_{js} = 0$ and $c_{jt} = M$ for $t = 1, \ldots, n$, where M is an arbitrarily large number. Set inventory holding costs h and demands D according to the *demand assignment cost* d_{ji} as follows:

$$\begin{array}{ccc} h_{1s}D_1 = d_{11} & \cdots & h_{ms}D_1 = d_{m1} \\ (h_{1s} + h_{11})D_2 = d_{12} & \cdots & (h_{ms} + h_{m1})D_2 = d_{m2} \\ \vdots & & \vdots \\ (h_{1s} + ... + h_{1,n-1})D_n = d_{1n} & \cdots & (h_{ms} + ... + h_{m,n-1})D_n = d_{mn} \end{array}$$

Or more generally,

$$(\sum_{\tau=s}^{t-1} h_{j\tau})D_t = d_{jt}, \; for \; t = 1, ..., n. \tag{11.A.1}$$

Further, set $D_s = 0$ and $I_0 = 0$. In the case where facility j cannot be used to satisfy the demand of point i, i.e., when $d_{ji} = \infty$ in a *UFL* instance, set d_{ji} to an arbitrarily large number *M*. Thus, so long as the holding cost h_{jt} is not restricted in sign, it is always possible to construct an instance of h_{jt} and D_t that satisfies all equations above. This and the above parameter setting allow us to construct an *UCAP-N* (with parameters c, h, K, D) instance given any *UFL* instance (with parameters d, f).

Now consider the solution of this specially constructed *UCAP-N* instance. From Theorem 11.1 (the non-splitting property) we know that there exists an optimal solution to *UCAP-N* where item i's demand in period t is produced in exactly one of the J_i facilities. Since $c_{js} = 0$ and $c_{jt} = M$ for $t = 1, \ldots, n$, production will only take place in period s at facility j. Demands in periods 1 to n are satisfied completely by inventory carried over from period s, i.e.,

$$x_{js} = \sum_{t=1}^{n} D_t \delta_{jt}, \forall j, \; and \; I_{jt} = \sum_{\tau=t+1}^{n} D_\tau \delta_{j\tau}, \forall j, t \tag{11.A.2}$$

Thus, the optimal solution to the *UCAP-N* instance has a corresponding *UFL* solution as follows:

(i) A production setup corresponds to facility opening, i.e.,

$$\chi_j = 1 \; if \; \sum_{t=s}^{n-1} \delta_{jt} \geq 1 \tag{11.A.3}$$

$$\chi_j = 0, \; Otherwise \tag{11.A.4}$$

(ii) Production setup costs correspond to facility opening cost f_j, for each $\chi_j = 1$, i.e.,

$$\sum_{t=s}^{n-1} \sum_{j=1}^{m} K_j \delta_{jt} = \sum_{j=1}^{m} f_j \chi_j \tag{11.A.5}$$

(iii) Total holding costs correspond to demand assignment costs (from eq. (11.A.1)(11.A.2)), i.e.,

$$\sum_{t=s}^{n-1}\sum_{j=1}^{m}(h_{jt}I_{jt}) = \sum_{t=s}^{n-1}\sum_{j=1}^{m}(h_{jt}(\sum_{\tau=t+1}^{n} D_\tau\delta_{j\tau}))$$

$$= \sum_{t=1}^{n}\sum_{j=1}^{m}((\sum_{\tau=s}^{t-1} h_{j\tau})D_t)\delta_t = \sum_{i=1}^{n}\sum_{j=1}^{m} d_{ji}y_{ji}$$

In summary, the optimal value of the specially constructed *UCAP-N* instance produces the optimal value of the *UFL* instance, i.e., in an optimal *UCAP-N* solution,

$$\sum_{t=s}^{n}\sum_{j=1}^{m}(c_{jt}x_{jt} + K_j\delta_{jt} + h_{jt}I_{jt}) = \sum_{i=1}^{n}\sum_{j=1}^{m} d_{ji}y_{ji} + \sum_{j=1}^{m} f_j\chi_j$$

Thus, if the optimal solution of the *UCAP-N* instance has a total cost less than κ, then there is a corresponding *UFL* solution that has a total cost less then κ (a *yes* answer to the decision problem). Otherwise, there is no assignment of demands to facilities with total cost lower than κ (a *no* answer).

Since *UFL* is known to be *NP-complete,* this proves that *UCAP-N* is also *NP-complete.* ◇

Proof of Corollary 11.3.

The proof for this corollary can be done by a simple observation as follows: if all D_t's are constants in equation (11.A.1) above, we can still construct an *UCAP-N* instance by finding h_{jt}'s. The rest remains the same as the standard case. ◇

Proof of Corollary 11.4.

To prove that when the setup cost K_{jt} is constant across facilities *UCAP-N* remains *NP-complete* we only need to show that *UFL* with constant facility opening cost f_j is *NP-complete.* We will show that the vertex cover (*VC)* problem can be polynomially reduced to *UFL.* For a given vertex cover instance, we construct a *UFL* instance as follows: associate each *edge* in *VC* to a demand point, and each *vertex* to a facility. When an edge is (not) incident to a vertex, the corresponding demand point can (not) be assigned to a facility (i.e., there is an edge (i, j) between the demand point and the facility in *UFL*). Set $d_{ji} = 0$ for all edges present and set facility opening cost f_j to a constant 1 for all facilities j. It is easy to verify that a solution to this specially constructed *UFL* instance corresponds to a solution to the original *VC* instance.

Therefore, *UFL* with constant facility opening cost is *NP-complete,* thus *UCAP-N* with constant setup cost is *NP-complete.* ◇

Proof of Theorem 11.7. The proof is similar to that of Theorem 11.2. It is therefore sufficient to outline the reduction from Uncapacitated Facility Location (*UFL*) to (P_i'). For a given instance of *UFL* problem we may construct a (P_i') instance as follows:

Given an UFL instance...	Construct a (P_i') instance such that
Demand points (i), i=1,...,n	Periods (t), t=1,...,n
Facilities (j), j=1,...,m	Facilities (j), j=1,...,m
Facility opening cost f_j j=1,...,m	Setup cost K_j, j=1,...,m
Demand assignment cost d_{ij}	Transportation cost $r_{jt}=d_{tj}$ ∀i,j,t
	Production cost $c_{j1}=0$ and $c_{jt}=m+1$ for t=2,...,n
	Holding cost $h_{jt}=0$, for j=1,...,m t=1,...,n
	Demands $D_t=1$

Consider the solution of this specially constructed (P_i') instance. The following are true:

1 From Theorem 11.1 (the non-splitting property) we know that there exists an optimal solution to (P_i') where item i's demand in period t is produced in exactly one of the J_i facilities. Since $c_{j1} = 0$ and $c_{jt} = m + 1$ for $t = 2, \ldots, n$, production will only take place in period 1 at facility j. Demands in periods 2 to n are satisfied completely by inventory carried over from period 1, i.e., production in period 1 at facility j is

$$x_{j1} = \sum_{t=2}^{n} D_t \delta_{jt}, \forall j,$$

and

$$I_{jt} = \sum_{\tau=t+1}^{n} D_\tau \delta_{j\tau}, \forall j, t$$

2 Total transportation costs correspond to demand assignment costs, i.e.,

$$\sum_{t=1}^{n} \sum_{j=1}^{m} r_{jt} b_{jt} = \sum_{i=1}^{n} \sum_{j=1}^{m} d_{ji} y_{ji}$$

3 Total setup costs correspond to facility opening costs

It is easy to verify that the optimal value of the specially constructed (P_i') instance produces an optimal value of the *UFL* instance. Thus, if the optimal solution of the (P_i') instance has a total cost less than κ, then there is a corresponding *UFL* solution that has a total cost less then κ (a *yes* answer to the decision problem). Otherwise, there is no assignment of demands to facilities with total cost lower than κ (a *no* answer).

Since *UFL* is known to be *NP-complete,* this proves that (P_i') is also *NP-complete.* ◇

References

Afentakis, P., B. Gavish and U. Karmarker (1984), "Computational Efficient Optimal Solutions to the Lot-Sizing Problems in Multistage Assembly Systems," *Management Science* 30, 222-39.

Arntzen, B.C., G.G. Brown, T.P. Harrison, and L.L. Trafton (1995), "Global Supply Chain Management at Digital Equipment Corporation." *Interfaces* 25(1) January-February 1995, 69-93.

Assad, A.A., (1980), "Models for Rail Transportation," *Transportation Research* 14A, 205-220.

Baker, K.R. (1993), "Requirements Planning" in *Logistics of Production and Inventory, Handbooks in Operations Research and Management Science, Volume 4,* ed. S.C. Graves, et al., Elsevier Science Publishers, Amsterdam, The Netherlands.

Bahl, H.C., Ritzman, L.P., and Gupta, J.N.D. (1987), "Determining Lot Sizes and Resource Requirements: A Review," *Operations Research* 35, 329-345.

Billington, P.J., J.O. McClain, and L.J. Thomas. (1983), "Mathematical Programming Approaches to Capacity-Constrained MRP Systems: Review, Formulation and Problem Reduction." *Management Science* 29(10), 1126-1141.

Blackburn, J., and R. Millen (1982), "Improved Heuristics for Multistage Requirements Planning Systems." *Management Science,* 28(1), 44-56.

Balakrishnan, A., and J. Geunes (2000), "Requirements Planning with Substitutions: Exploiting Bill-of-Materials Flexibility in Production Planning," *M&SOM* 2(2).

Burton, R.M., and Obel, B., (1984), *Designing Efficient Organizations,* North Holland, Amsterdam.

Carlson, R.C., and C.A. Yano (1983), "Safety Stocks in MRP - Systems with Emergency Setups for Components" *Management Science* 32(4), April, 403-412.

Cohen, M.A., and H.L. Lee. (1988), "Strategic Analysis of Integrated Production-Distribution Systems: Models and Methods." *Operations Research* 36(2), 216-228.

Cohen, M.A., and H.L. Lee. (1989), "Resource Deployment Analysis of Global Manufacturing and Distribution Networks." *J. of Manufacturing and Operations Management* 2, 81-104.

Davis, T. (1993), "Effective Supply Chain Management." *Sloan Management Review,* Summer, 35-46.

Dreyfus, S.E. and R. A. Wagner (1972), "On the Steiner Problem in Graphs," *Networks,* 1, 195-207.

Ertogral, K. and S.D. Wu (1999), "Auction Theoretic Coordination of Production Planning in the Supply Chain," *IIE Transactions, Special Issue on Design and Manufacturing: Decentralized Control of Manufacturing Systems,* forthcoming.

Federgruen, A. and M. Tzur (1981), "A Simple Forward Algorithm to Solve General Dynamic Lot Sizing Models with n Periods in O(n long n) or O(n) Time," *Management Science,* 27, 1-18.

Goyal, S.K. and Gunasekaran, A., (1990), "Multi-Stage Production-Inventory Systems," *European J. of Operational Research* 46, 1-20.

Graves, S.C., et al., (eds.) (1992), *Handbooks in Operations Research and Management Science, Volume 4, Logistics of Production and Inventory.* Amsterdam; Elsevier Science Publishers.

Guignard, M. and Kim, S. (1987), "Lagrangean Decomposition: A Model Yielding Stronger Lagrangean Bounds," *Mathematical Programming* 39, 215-228.

Gupta, S.M. and L Brennan. (1995), "MRP systems under supply and process uncertainty in an integrated shop floor control environment." *International J. of Production Research* 33(1), 205-220.

Hahm, J., and C. A. Yano. (1995a), "The Economic Lot and Delivery Scheduling Problem: the Common Cycle Case." *IIE Transactions* 27, 113-125.

Hahm, J., and C. A. Yano. (1995b), "The Economic Lot and Delivery Scheduling Problem: Models for Nested Schedules." *IIE Transactions* 27, 126-139.

Jörnsten, K.O., and Leistern, R. (1994), "Aggregation and Decomposition for Multi-Divisional Linear Programs," *European Journal of Operational Research* 72, 175-191.

Kennington, J.L. and R.V. Helgason (1980), *Algorithms for Network Programming,* Wiley-Interscience, New York.

Kimms, A. (1997), *Multi-Level Lot Sizing and Scheduling,* Physica-Verlag, Heidelberg, Germany.

Lee, H., P. Padmanabhan, and S. Whang (1995), "Information Distortion in a Supply Chain: The Bullwhip Effect." *Working Paper,* Stanford University, April, 1995.

Liebling, T. (1999), private communications.

Millar, H.H. and Yang, M. (1993), "An Application of Lagrangean Decomposition to the Capacitated Multi-Item Lot Sizing Problem," *Computers & Operations Research* 20, 409-420.

Sterman, J.D. (1989), "Modeling Managerial Behavior: Misperceptions of Feedback in a Dynamic Decision Making Experiment." *Management Science* 35(3).

Tayur, S., R. Ganeshan and M. Magazine, eds. (1999), *Quantitative Models for Supply Chain Management,* Kluwer's International Series, Kluwer Academic Publishers, Norwell, MA.

Tempelmeier, H. and M. Derstroff (1996), "A Lagrangean Based Heuristic for Dynamic Multilevel Multiitem Constrained Lotsizing with Setup Times," *Management Science* 42, 738-757.

Thizy, J.M., (1991), "Analysis of Lagrangian Decomposition for the Multi-Item Capacitated Lot-Sizing Problem," *INFOR* 29, 271-283.

Vidal, C.J. and M. Goetschalckx (1997), "Strategic Production-Distribution Models: A Critical Review with Emphasis on Global Supply Chain Models," *European J. of Operational Research* 98, 1-18.

Wagner, H.S. and T.M. Whitin (1958), "Dynamic Version of the Economic Log Sizing Model," *Management Science* 5, 89-96.

Wagelmans, A., S. Van Hoesel and A. Kolen (1992), "Economic Lot Sizing: An $O(n \ln n)$ Algorithm that Runs in Linear Time in the Wagner-Whitin Case," *Operations Research* 40, 145-155.

Chapter 12

AN OPTIMIZATION FRAMEWORK FOR EVALUATING LOGISTICS COSTS IN A GLOBAL SUPPLY CHAIN: AN APPLICATION TO THE COMMERCIAL AVIATION INDUSTRY

Amy Z. Zeng
Department of Management
Worcester Polytechnic Institute
Worcester, MA 01609
azeng@wpi.edu

Abstract This paper addresses logistics decisions for an aviation firm and its joint venture in Chengdu, China. The major logistics cost items for outsourcing aircraft engine components to the joint venture are identified and a number of transportation alternatives are examined based on cost effectiveness. Under the current structure of the company's global supply chain, we have evaluated five transportation options for moving materials inbound and outbound: all air, water-rail full container load, water-rail less than container load, water-truck full container load, and water-truck less than container load. A cost optimization model for each of these transportation modes using shipping quantity as a major decision variable is developed and the associated solution procedure is provided. These models provide useful guidelines for formulating sound transportation policies. Although the evaluation framework is established based on a particular company's global supply chain, it is applicable for many companies that are practicing global sourcing strategies.

Keywords: Global supply chain, Global sourcing, Logistics system, Transportation, Optimization

J. Geunes et al. (eds.),
Supply Chain Management: Models, Applications, and Research Directions, 317–341.

1. Introduction

Logistics processes comprise essential elements of fulfilling customers' orders. They form the critical loops of supply chains and oversee the flows of materials, information and cash. The costs associated with logistics activities normally consist of the following components: transportation, warehousing, order processing/customer service, administration, and inventory holding (e.g., Lambert et al. (1998); CMA Magazine, anonymous (1999)). Not surprisingly, total logistics costs represent a large portion of total supply chain costs, especially when the supply chain is extended to global markets. Previous studies have found that logistics costs have ranged from 4 to over 30 percent of sales (Ballou (1999)). As more organizations are outsourcing their products or services from all corners of the world, it becomes more critical for companies to understand and be able to estimate the various logistics cost components in order to assure their profit margin. This project analyzes the global logistics system of Company P in the US aviation industry (as requested by the company, we disguise the company's name), with emphasis on its transportation policies for moving materials between its headquarters and its joint venture located in Chengdu, China. In what follows, we will first provide a general background of the two companies and a description of the motivation for this research project. Then, we will explain the objectives, expectations and major results of this research effort.

1.1 Background

Company P is a leader in the design, manufacture and support of engines for commercial, military and general aviation aircraft, and space propulsion systems, as well as a pioneer in flight and technology. With an understanding of the nature of demand in today's competitive aviation industry, which is thrust at the lowest possible cost and highest level of reliability, the company strives to provide its customers with the services they need to focus on flying people and cargo safely around the world. In a business that shrinks the globe, Company P is truly worldwide: it has representatives in 76 cities in 47 nations. Its partnerships and joint ventures have reached to Asia and Europe and have kept the company at the forefront of flight.

Given the nature of aircraft engines, vertical integration is traditionally the dominant form of business structure in the aviation industry. However, potential market penetration and competition from foreign suppliers have led several big aircraft engine manufacturers in the U.S. to go global. With the decrease in U.S. military jet engine development in the early 1990's, international sourcing has taken a more strategic role. One example of such outsourcing strategy for Company P is a green-field joint venture located in the Sichuan province of China and established in 1996 through governmental arrangements (again, we will use Company C for this joint venture to disguise its identity). Company P supplied the initial capital investment and machinery for the project and continues to provide technical assistance and tooling for Company C. Manufacturing at the joint venture is divided into four major centers. These consist of welded sheet metal fabrication, major rotating part medium machining, simple machining and brazed assembly. These centers manufacture four types of engine parts: burner cans, pin disks, shrouds and shroud vane assemblies.

With a joint venture located in the Far East, Company P is increasingly concerned with the logistics costs associated with moving raw materials and finished products. The company and its joint venture have formed a vendor-required-material (VRM) relationship from the first day of operation of the joint venture. In this VRM relationship, Company P purchases the raw materials from one of its licensed suppliers and sends the materials through a preferred freight forwarder to the joint venture in China. It also holds the financial responsibility for shipping the material to and from Company C and paying for the value-added service provided by the joint venture. Adding to this heavy financial burden is the major transportation mode that is currently utilized by Company P: the raw materials as well as the finished products are all shipped by air between its headquarters and Chengdu, China, due to various reasons. Therefore, one of the major problems confronted by Company P is the estimation of the cost and profitability associated with the products outsourced to China.

1.2 Objective of This Study

The feasibility of outsourcing to China involves numerous issues, such as the internal operations of both companies, coordination between the two companies, Company P's partnerships with raw material suppliers, freight forwarders and end-customers, China's inland transportation infrastructure, transportation decisions, and import-export procedures, to name just a few. In this portion of the project, we have focused our attention and efforts on transportation policies. Specifically, we have identified major logistics cost components and classified them into different categories for the company to compile and collect data on in the future. We have also developed a decision framework with the aid of optimization models to evaluate the cost-effectiveness of various possible transportation methods. The minimum total logistics costs associated with each transportation mode, along with the shipping quantity, for materials transferred to and from Chengdu can be obtained and thus compared. A procedure for conducting sensitivity analysis is also provided based on the optimization models. It is necessary to mention that our intention in this project is not to point out which transportation mode is the best, since the final decision depends on numerous factors; instead, the goal is to provide methods for Company P to estimate the logistics costs resulting from global sourcing.

Since Company C is located in Chengdu, in the Sichuan province of China, which is an inland, mountainous and medium-sized city and far away from a major metropolitan area, the shipping route for Companies P and C consists of two major segments: outside of China and within China. After careful consultation with the two companies, we present the transportation routes in Company P's global supply chain in Figure 12.1. Figure 12.1 illustrates the possible transportation modes and routes for shipping both raw materials and finished goods. We see from Figure 12.1 that air can be used entirely between the two companies and that, if water is chosen for transport between Company P and Shanghai (a major port city in China and about 1,000 kilometers from Chengdu), then either truck or rail can be utilized to move materials between Shanghai and Chengdu. Therefore, three potential transportation modes, namely all air, water-rail combination, and water-truck combination, should be considered and compared. At present, the latter two options are not being practiced and there is no evaluation framework in place to help Company P make sound transportation decisions.

Thus, the major objectives of this study are three-fold and explained as follows. First, the major cost components involved in this global logistics systems need to be identified and documented for analysis. Although these cost items are derived based on this case study, we believe that they are excellent representatives of many global logistics systems, or at minimum serve as good references for many companies that are outsourcing materials in foreign countries. Second, optimization models that seek to minimize total logistics costs as a function of shipping quantity for each transportation mode are developed. Third, solution procedures for the cost optimization models are provided, which eventually yield the best combination of transportation modes for inbound and outbound movements. Finally, a framework for performing sensitivity analyses of in-transit time (or delivery lead time) and capital interest rate for the final mode selection is suggested and illustrated.

This chapter is organized in the following manner. In Section 2, a review of existing literature related to logistics cost analysis is given

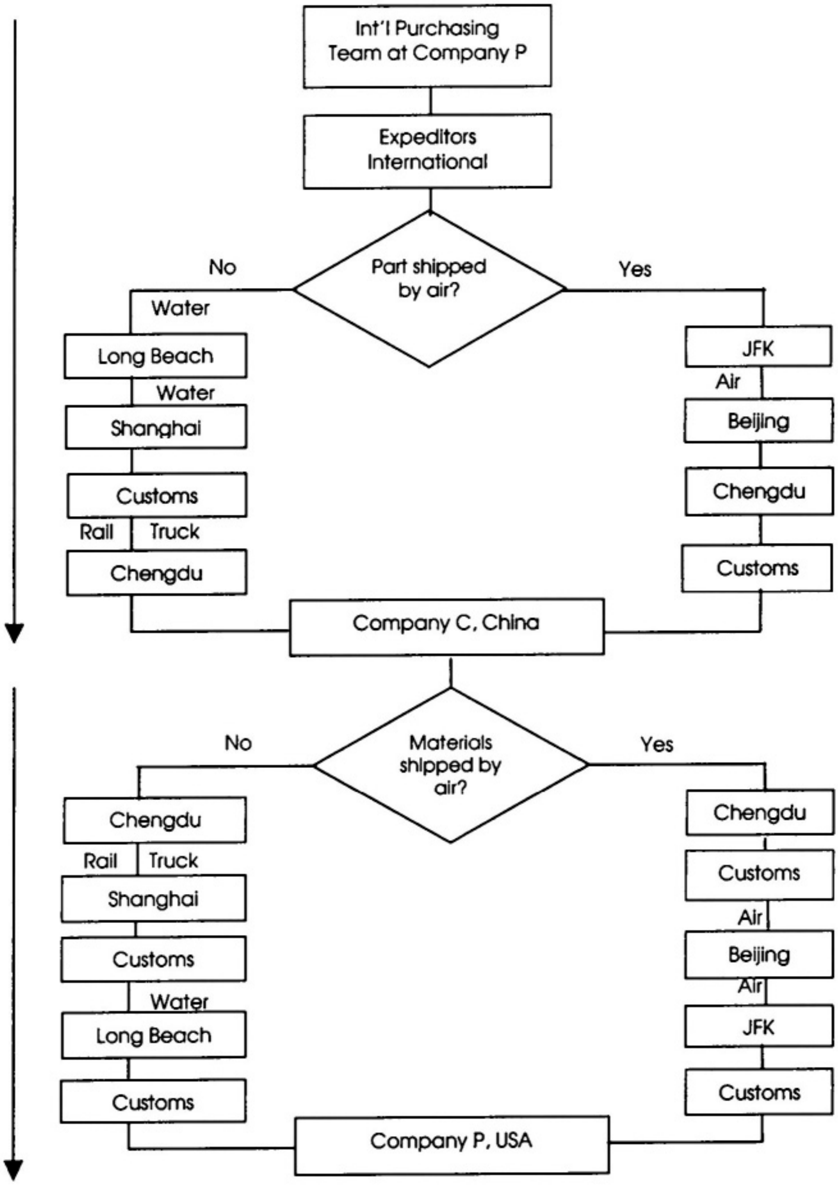


Figure 12.1. Transportation routes for the material flow

first, followed by a description of the cost components considered in Company P's global logistics system. In the last portion of this section, we present the general optimization model in detail. In Section 3, the development and analysis of the cost minimization models for all transportation modes, along with the solution procedures, are provided. Section 4 contains a numerical analysis of the current practice at Company P, sensitivity analysis, and a summary of managerial implications and discussion of the limitations of the mathematical models. Finally, concluding remarks are given in Section 5.

2. General Modeling Approach

2.1 Total Logistics Costs: Literature Review

Before the general model for this research is presented, we first provide some theoretical background based on existing literature. We divide the literature that reports study results on logistics cost into two streams. One stream focuses on strategic elements of logistics, and the other deals with optimal logistics decisions. A large amount of literature exists discussing the strategic role of logistics in creating value and its relationship to a company's financial performance. As reported by Richardson (1995), logistics controls a significant amount of assets and has direct impact on cash flow and the bottom line, adds value through continuous productivity and service improvements, and possesses a strong relationship with a firm's customer service level and revenues. As global sourcing is rapidly arising as a prerequisite for competing in today's markets, capturing and evaluating the logistics costs involved in a global supply chain appears as increasingly critical and important as uncovering the strategic benefits (Fagan (1991)). The greatest challenge faced by the logistics professional stems from the nature of the global sourcing process: it is time-consuming, complex and volatile (McGowan (1997)). Numerous factors can drive up logistics costs substantially, which may offset the benefits of doing business with international suppliers. One way to account for the logistics costs is presented by van Damme and van der Zon (1999), who provide an activity-based costing approach for analyzing financial information and helping top management make logistics decisions. Maltz and Ellram (1997) identified those logistics activities that affect outsourcing decisions and presented a ten-step procedure for comparing make or buy alternatives using total cost relationships.

The major trend of the second stream focuses on examining system costs, which include transportation costs in conjunction with inventory and purchasing expenses. Examples of studies restricted to a domestic logistics system include Lee (1986); Russell and Krajewski (1991); Tersine and Barman (1991); Bertazzi et al. (1995), and Tyworth and Zeng (1998), where the optimal order/purchase quantity is derived based on minimizing the system cost for a particular transportation mode. Another characteristic of these efforts is the inclusion of the effect of freight or quantity discounts on the ordering quantity and the associated total cost. Demand patterns considered in these research efforts are assumed to be deterministic, except in the study of Tyworth and Zeng (1998), and issues in vehicle routing, scheduling and consolidation are ignored except in the article of Bertazzi et al. (1995). Few research efforts have been devoted to evaluating transportation policies in the context of global supply chains. Vidal and Goetschalckx (1997) present a framework and point out the future research opportunities for modeling international production-distribution chains. There are two studies on logistics cost that bear some resemblance to this project. One effort given by Liao (1997) focuses on the freight cost structure for three major shipping modes and the cost of locating manufacturing facilities in a foreign country. The only international factor included in the total cost model is that of duties and taxes. The impact of holding pipeline inventory, a major cost driver in any global logistics systems, and the order processing cost are ignored. In addition, the solution algorithm is limited to a domestic-system environment. The other study completed by Fera (1998) attempts to identify and then classify the relevant cost factors for evaluating the feasibility of the international outsourcing strategy of Company P. Although a comprehensive list of recurring and non-recurring cost items associated with international sourcing is provided, minimization of the total cost is not attempted, nor is the evaluation of the various transportation alternatives.

Our project modifies the general cost drivers developed by Maltz and Ellram (1997) to accommodate Company P's situation and aims at developing cost minimization models that account for the major cost factors in a global logistics system and at assessing the economic performance of the available transportation choices. The theoretical results from the optimization framework will provide helpful guidelines for the international purchasing group at Company P to formulate sound transportation policies. In what follows, the cost components will be presented and discussed first, followed by the general optimization model and its assumptions.

2.2 Logistics Cost Components Considered in This Study

The greatest difficulty associated with this study is the identification of the major cost components. After a couple of months of analysis of Company P's practices, we have classified the key cost items into the following six categories: transportation, inventory holding, administration, customs charges, risk, and handling and packaging, as shown in Figure 12.2 (a detailed discussion of these cost items is provided in a project report prepared for the company, which is available from the author upon request). We can see from Figure 12.2 that each category consists of a few cost items and the dimension of each cost item is also given. It is necessary to mention that although duties/tariffs are commonly incurred in international trade, they do not apply to this case, because the joint venture in China does not import anything for sale within China and Company P is a participant in a protected industry. As indicated by the dimensions, cost items fall into one of the following categories: weight-based ($/kg), value-based (%), frequency-based ($/shipment), or time-based ($/$/year). While previous research has suggested some guidelines for classifying logistics costs, such as activity-based or occurrence-based, our classification is primarily driven by the availability of the company's data, the possibility to derive the needed information based on the given data, and the format we used to construct the optimization models. We have shown the total cost structure to related personnel at Company P and received agreement. We next move ahead to the cost minimization model.

2.3 The Optimization Model

Apparently a large number of cost items inevitably accrue when moving materials along a global supply chain. Figure 12.2 contains notation for each type of cost component and subcategories within each cost category are denoted using numerical subscripts. For instance, C_{h_1} refers to the fractional charge of holding pipeline inventory per dollar per year. In addition, the following notation will be used throughout this chapter.

D = Total annual demand (units/year)

v_r = Value of the raw material per unit ($/unit)

v_f = Value of the finished product per unit ($/unit)

w_r = Unit weight of raw material (kg/unit)

w_f = Unit weight of finished part (kg/unit)

t_p = Total in-transit time from one destination to the other (years).

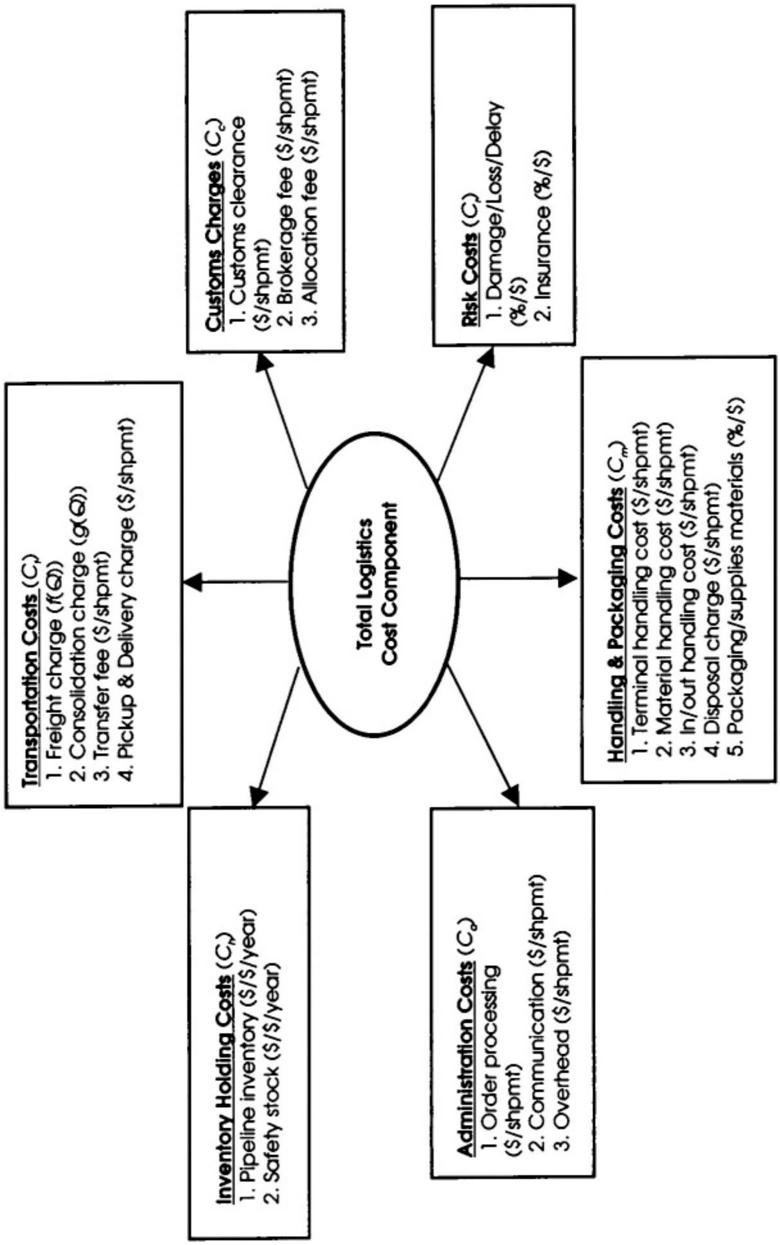


Figure 12.2. Cost components in the global logistics system

In addition, the decision variables are:

Q_f = Shipment size of the finished parts per year (kg/shipment)
Q_r = Shipment size of the raw material per year (kg/shipment).

Before we present the mathematical models, it is necessary to point out the scenarios we will examine and some general assumptions. We will restrict our attention to the transportation issues associated with the joint venture in China; hence, the costs associated with purchasing the raw materials and manufacturing at Company C are not considered. Additionally, the assumptions used to develop the models are explained as follows.

(1) The demand for products outsourced to Company C is deterministic, and the associated demand for the raw materials can be derived based on the information on the finished items. As Company P has established long-term relationships with its customers, the demand for the parts to be used in building aircraft engines remains fairly stable.

(2) Since Company C supplies more than one finished engine component, aggregate values of the product characteristics, such as the demand, weight, volume, and value for both raw materials and finished goods, will be utilized. This assumption is reasonable since materials at both their initial and final stages have well-defined dimensions.

(3) Manufacturing intervals accommodate shipping frequencies, i.e., Company C always has the quantities of finished goods ready for shipment for any value of shipment size derived from the cost minimization model. The same assumption holds for the raw materials suppliers.

Note that since the U.S. dollar is used as the main currency and Company P pays all the logistics costs, the currency exchange rate is not an issue in this study.

In what follows, the optimization models for finding the optimal shipment size of raw materials from the parent company to the overseas supplier and the shipment size of the finished products in the opposite direction will be developed and discussed. Note that since the cost components are symmetric for both directions, we will ignore the subscripts (r, f) in the models. Given the cost categories listed in Figure 12.2, the general cost minimization model can be given as follows:

minimize: Expected Total Annual Cost (ETAC) =
Transportation + Holding + Customs + Administration
+ Risk + Material Handling Costs

Each cost component is calculated on a yearly basis (i.e., $/year) and the formulas are given as follows:

$$\begin{aligned}
\text{Transportation cost} &: \quad C_t = f(Q) + g(Q) + \frac{Dw}{Q}\left(C_{t_3} + C_{t_4}\right) \\
\text{Inventory holding cost} &: \quad C_h = (Dv)C_{h_1}t_p + Q(v/w)C_{h_2} \\
\text{Customs charge} &: \quad C_c = \frac{Dw}{Q}\sum_{i=1}^{3} C_{c_i} \\
\text{Administration cost} &: \quad C_a = \frac{Dw}{Q}\sum_{i=1}^{3} C_{a_i} \\
\text{Risk cost} &: \quad C_r = (Dv)\left(C_{r_1} + C_{r_2}\right) \\
\text{Material handling cost} &: \quad C_m = \frac{Dw}{Q}\sum_{i=1}^{4} C_{m_i} + (Dv)C_{m_5}
\end{aligned}$$

Most of the above cost functions are straightforward based on the dimension of each of the cost item, but the inventory holding cost deserves some explanation and clarification. The first term indicates the annual cost of carrying pipeline inventories (where t_p denotes the time inventory spends in the pipeline. This item encompasses the total time spent on a series of events that take place when moving materials. The moving process typically begins with consolidation, and involves transferring the consolidated goods to the airport/sea port by rail or truck, storage in warehouses, loading, actual transit, unloading, customs clearance, transfer to the destination, and finally ends with receiving.) The second term calculates the safety stock holding cost. It is necessary to mention that although the annual demand is considered deterministic, given the uncertainties involved in a global supply chain, we assume that one shipment size is always held as a buffer stock in case of emergencies or discrepancies occurring between demand and supply. This assumption is also being practiced at the two companies. Note also that the cycle stock holding cost is not included, as it is incurred on the shop floor of the joint venture and not paid by Company P (all cost components in the model are paid by Company P). Letting

$$\theta_1 = C_{t_3} + C_{t_4} + \sum_{i=1}^{3} C_{a_i} + \sum_{i=1}^{4} C_{m_i}$$

and

$$\theta_2 = C_{r_1} + C_{r_2} + C_{m_5}$$

then the total cost function can be written as

$$\text{ETAC} = \frac{Dw}{Q}\theta_1 + Q(v/w)C_{h_2} + (f(Q) + g(Q)) + (Dv)(\theta_2 + C_{h_1}t_p).$$

Notice that two cost components, namely the freight cost $f(Q)$, and the consolidation cost $g(Q)$, are expressed as general functions. The consolidation cost is incurred when materials are crated in order to accommodate carriers' needs. The form of these functions will depend on the transportation mode under consideration and the rate scheme charged by the carriers. In the next section, we will study the explicit total cost function for each of the transportation options considered feasible by both Company P and Company C.

3. Analysis of Transportation Alternatives

The current transportation structure of the company's global supply chain as shown in Figure 12.1 suggests that the combinations of water, truck and rail are possible shipping approaches in addition to air. When these options are considered, both full container load (FCL) and less-than container load (LCL) options should be examined. As a result, a total of five potential transportation alternatives, namely water-rail FCL, water-rail LCL, water-truck FCL, water-truck LCL, and air, become potential candidates for moving raw materials and finished goods. Note that if the ocean is used, the materials must be transferred from Company P's headquarters to Long Beach, California by either rail or truck. This portion of cost within the U.S. is taken into account as follows. The total in-transit time will include three major segments: the travel time across the U.S., the transit time between the U.S. and China, and the transit time within China. We assume that the same mode is used when goods are moved within each of the two countries (i.e., if truck is used for transporting within China, then the same is used within the U.S.) and that the same freight rate is used for both countries due to the unavailability of data from the two countries. In the following subsections, we develop cost minimization models for all five shipping alternatives.

3.1 Water-Rail Combination

In this scenario, we will examine the effectiveness of using FCL or LCL for both water and rail alternatives.

3.1.1 Water-Rail FCL. This option implies that both water and rail FCL are employed. In this case, the FCL cost structure for water using Landbridge (a service in which foreign cargo crosses a country en route to another country; see Lambert et al. (1998)) is a fixed, one-time charge for each shipment. The cost charged for using a full load in rail depends on the container volume, not the shipping size. The freight charge for the associated consolidation fee is also a constant for each shipment. Let the freight charge of ocean and rail and the consolidation fee be denoted as C_{t_w}, C_{t_r}, and C_{t_2}, respectively. Hence, the annual freight cost i s calculated a $C_{t_1} = (Dw/Q)(C_{t_w} + C_{t_r})$. e total cost function is given as follows:

$$\begin{aligned} \mathrm{ETAC}_{\mathrm{WR-FCL}} &= \frac{Dw}{Q}\left(\theta_1 + C_{t_w} + C_{t_r} + C_{t_2}\right) \\ &\quad + Q(v/w)C_{h_2} + (Dv)\left(\theta_2 + C_{h_1} t_p\right). \qquad (12.1) \end{aligned}$$

The structure of the cost function (12.1) is similar to that of the classic economic order quantity (EOQ) function (Silver et al. (1998)). Therefore, it is straightforward to show that the minimum cost occurs when the shipment size equals

$$Q^*_{\mathrm{WR-FCL}} = \sqrt{\frac{(Dw)\left(\theta_1 + C_{t_w} + C_{t_r} + C_{t_2}\right)}{(v/w)C_{h_2}}}. \qquad (12.2)$$

It is necessary to mention that since the full container load size is usually very large and the products considered in this study are of high value and low demand, the likelihood for the optimal shipment size in (12.2) to be greater than the full container load is almost zero. Hence, it is reasonable to assume that the optimization problem shown in (12.1) is unconstrained. Substituting the shipping quantity in (12.2) into the cost function (12.1) yields the minimum total cost in this scenario as

$$\mathrm{ETAC}^*_{\mathrm{WR-FCL}} = 2\sqrt{(Dv)\left(\theta_1 + C_{t_w} + C_{t_r} + C_{t_2}\right)C_{h_2}} + (Dv)\left(\theta_2 + C_{h_1} t_p\right).$$

3.1.2 Water-Rail LCL. When using the LCL option, the freight charge for water is calculated based on either the weight or the cubic size (measured in cubic meters, CBM) of the material to be shipped.

Since the two parameters have a linear relationship for a standard, well-shaped raw material and finished item, i.e., the rate expressed in \$/CBM can be converted to \$/kg, or vice versa, we continue to use \$/kg as the dimension for the ocean rate. The freight rate for rail and associated consolidation fee, however, are dependent on the container load size; letting L_j denote the container load size for container j, the rate structure can usually be displayed in the following format:

Load size breakpoints	Rail freight rate (LCL) (\$/container)	Consolidation fee (\$/container)
$0 < Q \leq L_1$	$C_{t_r}^{(1)}$	$C_{t_2}^{(1)}$
$\vdots$	$\vdots$	$\vdots$
$L_{j-1} < Q \leq L_j$	$C_{t_r}^{(j)}$	$C_{t_2}^{(j)}$
$\vdots$	$\vdots$	$\vdots$
$L_{m-1} < Q \leq L_m$	$C_{t_r}^{(m)}$	$C_{t_2}^{(m)}$

It is important to mention that the container volume is much smaller than that used in the FCL case and that the charge is not proportional to the shipment weight, i.e., a fixed amount of cost is incurred for using a container even if the container is not filled up. The $C_{t_r}^{(j)}$ and $C_{t_2}^{(j)}$ values are increasing at a decreasing rate with container size. This cost structure is similar to an all-units discount structure, except that here we have an *increasing step function* cost structure in the total quantity shipped. However, it is straightforward to show that a similar procedure for finding the optimal shipment quantity holds as for the all-units discount case (see Silver et al. (1998)). We first find the EOQ value that is realizable (falls within the proper quantity interval) starting from the highest discount level and then check the cost at all *lower* breakpoints. The reason for this is that, given the cost structure, we may be able to save on both shipping and holding costs by reducing the EOQ at lower discount breakpoint levels. Increasing to a higher breakpoint above the highest realizable EOQ increases holding costs, while either maintaining or increasing the per-shipment fixed cost. Note that L_m denotes the largest container and that the rail freight rate is not distance-based, rather, it is dependent solely upon weight.

We assume that in each option, only one container of each size is utilized. This assumption is reasonable as both the raw materials and finished goods are neither bulky nor heavy. The following solution algorithm is presented to determine the best shipment quantity.

Step 1. Calculate the shipment size at each LCL freight rate using the following formula

$$Q_j = \sqrt{\frac{(Dw)\left(\theta_1 + C_{t_r}^{(j)} + C_{t_2}^{(j)}\right)}{(v/w)C_{h_2}}}, \qquad (12.3)$$

and denote the shipment size that applies to both the freight rate and consolidation charge by Q_v. If $L_{j-1} < Q_v \leq L_j$, then this order quantity is realizable for the appropriate discount range. If so, calculate the associated total cost using

$$\text{ETAC}_{Q_v} = TC_Q + D\left(v\theta_2 + wC_{t_w}\right) + D(vC_{h_1}t_p), \qquad (12.4)$$

where

$$\begin{aligned} TC_Q &= \frac{Dw}{Q_v}\left(\theta_1 + C_{t_r}^{(v)} + C_{t_2}^{(v)}\right) + Q_v(v/w)C_{h_2} \\ &= 2\sqrt{(Dv)\left(\theta_1 + C_{t_r}^{(v)} + C_{t_2}^{(v)}\right)C_{h_2}}, \qquad (12.5) \end{aligned}$$

and the values $(C_{t_r}^{(v)}, C_{t_2}^{(v)})$ correspond to the weight range where the Q_v is located. Let Q_v^* and TC_Q^* denote the values of (12.3) and (12.5) at the realizable order quantity at the highest discount level.

Step 2. Let L_Q be the largest container load size less than Q_v^*. As the last two terms in (12.4) are independent of the shipping weight, only the first portion of the total cost needs to be calculated as follows:

$$TC_L = \frac{Dw}{L_Q}\left(\theta_1 + C_{t_r}^{(v)} + C_{t_2}^{(v)}\right) + L_Q(v/w)C_{h_2}. \qquad (12.6)$$

Repeat the calculation of Equation (12.6) at all discount breakpoints lower than L_Q and let TC_L^* denote the breakpoint value that gives the minimum value of (12.6).

Step 3. Choose the shipment amount that yields

$$TC^* = \min\left(TC_Q^*, TC_L^*\right).$$

The freight rate scheme used in the LCL case implies that using EOQ formula in (12.3) does not necessarily lead to a valid shipping weight for each pair of freight rate and consolidation charge. As a fixed freight rate and consolidation fee are charged for a range of shipping weight, a certain amount of savings in freight and consolidation costs can be realized if the range limit is used. Hence, it is necessary to compare the calculated quantity and the quantities at lower price breakpoints.

3.2 Water-Truck Combination

Truck is another alternative for transporting the materials between Shanghai and Chengdu. Analogous to rail, this option has two scenarios, namely FCL and LCL.

3.2.1 Water-Truck FCL. In this scenario, an FCL rate will be used for both water and truck. For water, the rate is a constant as mentioned in the preceding section; for truck FCL, the freight rate per shipment is also a constant and is denoted as C_{t_k}. In addition, the consolidation charge is measured on a per-shipment basis. The analysis follows in a similar manner to that in the water-rail FCL case. Letting the fixed FCL charge rate ($/shipment) be C_{t_w} and C_{t_k} for ocean and truck, respectively, the optimal shipment size can be obtained as

$$Q^*_{\mathrm{WK-FCL}} = \sqrt{\frac{(Dw)\,(\theta_1 + C_{t_w} + C_{t_k} + C_{m_5})}{(v/w)C_{h_2}}}.$$

3.2.2 Water-Truck LCL. This option is similar to the water-rail-LCL alternative, except that the charge rate for truck will be different from the rail rate. The dimension of the ocean freight rate remains in dollars per kilogram ($/kg), and the charging scheme for using truck is based on the truck container size, which is weight-based and independent of the distance according to the operational procedure at Company P. Note that in general, LCL rates depend on both weight and distance, but this is not the case for these two companies. Hence, the solution procedure presented in Section 3.1.2 remains the same for this scenario and will not be repeated here.

3.3 Air

Both raw materials and finished products can be shipped by air for the entire route. Company P normally uses a preferred freight forwarder for all shipments, whose freight rate scheme is described in table 12.1.

In this scheme, both the freight rate and the consolidation fee are charged in dollars per kilogram, which is different from the water-rail LCL case, and the their values are decreasing in volume and correspond to a *decreasing* step function structure. This scheme is equivalent to a fixed annual contract price for air freight plus a consolidation fee that depends on the shipment quantity (and, hence, frequency). A lower value of Q implies more frequent shipping and should, therefore, involve a higher annual cost. Because of the dimension of the structures above,

Table 12.1. Freight rate scheme for air shipments

Weight breakpoints	Air freight rate ($/kg)	Consolidation fee ($/kg)
$0 < Q \le W_1$	$C_{t_r}^{(1)}$	$C_{t_2}^{(1)}$
$\vdots$	$\vdots$	$\vdots$
$W_{j-1} < Q \le W_j$	$C_{t_r}^{(j)}$	$C_{t_2}^{(j)}$
$\vdots$	$\vdots$	$\vdots$
$W_{m-1} < Q \le W_m$	$C_{t_r}^{(m)}$	$C_{t_2}^{(m)}$

the annual freight cost and consolidation charge is computed as

$$f(Q) + g(Q) = \frac{Dw}{Q}\left(C_t^{(j)} + C_{t_2}^{(j)}\right)$$

for a shipment size Q_j. Furthermore, since a maximum allowable weight for each shipment is normally imposed for air shipping, it is important to incorporate this constraint into decision-making. Let this shipping limit be denoted as W_m (in kg, note that the W_m here is much smaller than the largest containers used in rail LCL and truck LCL, respectively). Thus, the optimization problem for finding the best air shipping quantity can be expressed as

$$\text{minimize ETAC}_{\text{air}} = \frac{Dw}{Q}\theta_1 + Q(v/w)C_{h_2} + (Dv)(\theta_2 + C_{h_1}t_p) + (f(Q) + g(Q)) \quad (12.7)$$

subject to

$$Q \le W_m.$$

It is straightforward to show that in this case, since both freight cost and consolidation charge have a *decreasing* step function structure, the same approach used in the standard all-units discounting scheme can be used, i.e., find the realizable EOQ value at the highest discount level, and check the cost at all higher discount breakpoints. However, the EOQ formula used in this problem is slightly different from that established in the literature. Since the value paid by the buying organization (Company P) excludes the freight and consolidation charges, the EOQ is independent of these two parameters and is calculated as follows:

$$Q_{\text{air}} = \sqrt{\frac{(Dw)\theta_1}{(v/w)C_{h_2}}}.$$

Therefore, once the value of Q_{air} is obtained, the freight and consolidation rate pair in the scheme that applies to Q_{air} can be determined. Then the associated total cost of using Q_{air} needs to be compared with the cost at all other freight and consolidation rate pairs at higher discount breakpoints.

Each of the prior sections was devoted to a particular transportation mode combination, in which the cost model and solution procedure are provided. Here the rate structure for freight and consolidation cost is the key for finding the solution algorithm. The optimal order quantity with all-unit discounts established in the literature serves as a good basis, but does not always apply to various cost structures in reality. In our study, we have extended the existing solution procedures to incorporate the cost schemes of not only the freight rates, but also consolidation rates.

4. Numerical Examples and Sensitivity Analysis

4.1 Computational examples

To demonstrate the effectiveness of the mathematical models presented in the preceding section, we have conducted a few numerical studies using the data and current practices provided by the companies. For the protection of confidential information, we will not use the freight forwarders' names. Also most of the data items are slightly modified, some of which are combined into one fixed parameter (such as θ_1 and θ_2). As a result, the fixed as well as the variable input parameters are summarized in Tables 12.2 and 12.3, respectively. One important parameter in Table 12.3, namely the in-transit time (t_p), deserves some attention. As defined before, this item encompasses the total time spent on a series of events that take place when moving materials. Due to the large number of events involved, the uncertainty associated with the in-transit time is normally very high, which is revealed by the possible delay time listed in Table 12.3. To capture the risk level involved in each transportation mode, we use the following percentages for computing the loss and damage cost associated with water-rail, water-truck and air, respectively: 1%, 2%, and 0.5% of total values shipped. The freight rate schemes for air, truck and rail for both directions are listed in Table 12.4. We will refer to the situation in which all parameters take the values indicated in Tables 12.2 and 12.3 as the base case.

The optimal shipping quantity, in kilograms per shipment, and the minimum total annual logistics cost for the base case are obtained using the solution algorithms provided in the preceding section and are reported in Table 12.5. We can observe from Table 12.5 that

Table 12.2. Fixed Parameters in the Numerical Example

Parameter			*Value (US to China)*	*Value (China to US)*
D		(unit/year)	2,900	2,900
v		(\$/unit)	1,000	3,000
w		(kg/unit)	31	29
C_{h_1}		(\$/\$/year)	0.12	0.10
C_{h_2}		(\$/\$/year)	0.10	0.10
C_{t_w}	FCL	(\$/shipment)	2,947	2,947
C_{t_w}	LCL	(\$/kg)	0.604	0.968
C_{t_r}	FCL	(\$/shipment)	1,960	1,960
C_{t_k}	FCL	(\$/shipment)	320	320

Table 12.3. Variable Parameters in the Numerical Example

Parameter			*US to China*			*China to US*		
			Air	Water-Rail	Water-Truck	Air	Water-Rail	Water-Truck
C_{t_2}	(FCL)		40	100	40	40	100	40
θ_1	(\$/shipment)		290	1,795	1,475	395	1,835	1,808
θ_2	(%)		3	3.5	4.5	3.3	4	4.8
t_p	(day)	Mean	5	50	33	5	47	36
t_p	(day)	Delay	3	20	18	5	22	17

Table 12.4. Rate Schemes

(1) Air, JFK to Chengdu								
Shipment weight	(kg)	50	100	150	300	550	1,000	2,000
Freight rate	(\$/kg)	4.65	4.10	3.25	3.10	2.95	2.80	2.75
Consolidation, C_{t_2}	(\$/kg)	0.80	0.70	0.65	0.60	0.50	0.45	0.40
Air, Chengdu to JFK								
Shipment weight	(kg)	50	100	150	300	550	1,000	2,000
Freight rate	(\$/kg)	5.60	5.20	4.80	4.70	4.60	4.40	4.35
Consolidation, C_{t_2}	(\$/kg)	0.80	0.70	0.65	0.60	0.50	0.45	0.40

(2) Rail LCL, Chengdu-Shanghai				
Container size		1,000	5,000	10,000
Freight rate	(\$/container)	200	600	900
Consolidation, C_{t_2}	(\$/container)	30	90	120
(3) Truck LCL, Chengdu-Shanghai				
Container size		3,000	5,000	8,000
Freight rate	(\$/container)	120	200	280
Consolidation, C_{t_2}	(\$/container)	10	15	20

Table 12.5. The Optimal Shipping Quantity and Total Logistics Cost for the Five Transportation Modes: the Base Case

	US to China		*China to US*	
Mode	Quantity (kg)	ETAC ($/yr)	Quantity (kg)	ETAC ($/yr)
W-R FCL	13,768	237,999	7,458	614,333
W-R LCL	8,857	260,615	4,531	635,175
W-K FCL	11,544	236,442	6,449	636,826
W-K LCL	7,033	261,639	4,055	668,722
All Air	2,000	394,277	1,792	735,568

(1) for materials moving from Company P to the joint venture, using the water-truck FCL shipping mode yields the lowest total cost ($236,442), and the water-rail FCL performs the best for the other direction ($614,333); and

(2) the maximum allowable weight $(W_m = 2000)$ is used for air shipping from the parent company to the joint venture.

Considering the total cost of both directions, we have calculated the sum of the costs incurred for both directions and presented the results in a matrix format as seen in Table 12.6. The numbers on the diagonal indicate the total cost using the same transportation mode for both directions. Table 12.6 shows that using air to move materials, although reducing the holding cost of pipeline inventory, is the most expensive method for the case considered due to its high freight cost. Table 12.6 also pinpoints the combined best transportation modes, which are water-truck FCL from the U.S. to China and then water-rail FCL for the way back, together leading to savings of $279,233 (= $1,130,008 – $850,775) compared to the cost of using all air. In summary, the mathematical models we have proposed not only identify the shipping quantity and associated logistics costs for a given transportation mode combination, but also help evaluate the various combinations of the transportation modes based on their cost effectiveness.

4.2 Sensitivity analysis

A quick comparison of the transportation modes considered in this study reveals their immediate advantages and disadvantages. For example, air shipping provides the most reliable and fastest service, which is reflected in the holding cost of pipeline inventories and risk charges. However, the high cost of transportation tends to offset the benefits. The other options enjoy lower transportation cost but involve high degrees

Table 12.6. The Minimum Total Logistics Cost of Materials Flow between Company P and China Using Various Transportation Modes: the Base Case (all numbers are expressed in thousands)

			China to US				
			W-R FCL	W-R LCL	W-K FCL	W-K LCL	All Air
			614.3	635.2	636.8	668.7	735,568
US to China	W-R FCL	238.0	852.3	873.2	874.8	906.7	941.8
	W-R LCL	260.6	874.9	895.8	897.4	929.3	964.4
	W-K FCL	236.4	**850.7**	871.6	873.3	905.2	940.2
	W-K LCL	261.6	876.0	896.8	898.5	930.4	965.4
	All Air	394.4	974.6	995.5	997.1	1,029.0	*1,130.0*

of variability and risk. Therefore, it would be worthwhile to compare their performance in various situations.

We rely on varying two parameters to conduct the sensitivity of the mode preference, namely the in-transit time and the inventory holding cost rate for the following reasons. First, based on our analysis of the company's data and previous practices, we have noticed that the uncertainty involved in the in-transit time of using ocean, rail and truck can be very high, and we have compiled data for both average and possible delay time shown in Table 12.3. Second, since the percentage holding cost rate depends on the opportunity loss of capital, which has a strong relationship with the interest and inflation rate, it is reasonable to use a range, rather than a fixed number, to describe the holding rate. Finally, we assume that the holding rate of pipeline inventories is more influential than that of the safety stock due to the amount of the pipeline inventories held in a global supply chain. Hence, we fix the safety stock holding rate but vary the pipeline holding rate. Moreover, we use the total cost of air shipping as a benchmark to evaluate the performance of other shipping alternatives.

Consequently, we vary the in-transit time, t_p, and the holding rate, C_{h_i}, as follows:

$$t_p = \text{base value (in days)} + i, \qquad i = 1, 2, \ldots, 7$$

and

$$C_{h_1} = \text{base value (percentage)} + i/100, \qquad i = 1, 2, \ldots, 7$$

for both shipping directions. These values generate 49 (= 7 × 7) combinations for each shipping direction. The cost values for all 49 problems were calculated and the optimal total cost of both directions for each

Table 12.7. Sensitivity Analysis Using Air as a Benchmark (Minimum annual cost of using air is $1,130,008)

Holding Rate Increase (%)	In-transit Time Increase (days) 1	2	3	4	5	6	7
1	850,775	854,112	857,449	860,786	864,123	867,460	870,797
2	864,600	868,254	871,909	875,564	879,219	882,873	886,528
3	878,424	882,397	886,369	890,342	894,315	898,287	902,260
4	892,249	896,539	900,830	905,120	909,410	913,701	917,991
5	906,073	910,682	915,290	919,898	924,506	929,115	933,723
6	919,898	924,824	929,750	934,676	939,602	944,528	949,454
7	933,723	938,967	944,210	949,454	954,698	959,942	965,186

problem was obtained in the following manner:

$$\text{Total Cost} = \text{min. cost}\ \{\text{W-R FCL; W-R LCL; W-K FCL; W-K LCL}\}_{\text{US}\rightarrow\text{Ch}} + \text{min. cost}\ \{\text{W-R FCL; W-R LCL; W-K FCL; W-K LCL}\}_{\text{Ch}\rightarrow\text{US}}.$$

The results are reported in a matrix shown in Table 12.7. As indicated by Table 12.7, even if the in-transit time of using combinations of water, rail and truck is extended by one week and the holding rate is increased by 7 percent, air remains the most expensive mode of transporting goods (Recall that the total cost of using air is $1,130,008 per year). The savings can range from $279,233 per year in the base case to $164,822 per year under worse conditions. Although we didn't consider the performance of water, truck and rail combinations in worse situations, this example illustrates the procedure for conducting sensitivity analysis of various combinations of the values. The matrix format, as presented in Table 12.7, is useful for comparing the performance of the transportation modes in different scenarios.

4.3 Discussion of model limitations and managerial implications

In the previous sections, we presented a set of mathematical models with the objectives of minimizing the total logistics costs and identifying the shipping size for various transportation alternatives, along with their solution algorithms. The numerical analysis has demonstrated that the models can be used for assessing the cost-effectiveness of various available transportation mode combinations, which will help the international purchasing group at Company P and its joint venture in China to make sound logistics decisions.

It is important to point out some of the limitations underlying the models. The first limitation has to do with the demand pattern, which is assumed to be level over time. This is valid only if both companies coordinate their ordering and manufacturing decisions. Efficient coordination can be only achieved through fast and accurate information sharing and communication, quick transfer of MRP schedules, accurate demand forecasting, and close relationships with raw materials suppliers. While these issues are beyond the scope of this portion of the project, they provide a foundation for the mathematical models.

The second limitation is that the effect of payment frequency is ignored in the models. As pointed out before, Company P is the major party responsible for most of the costs. Therefore, it is concerned with not only the magnitude of the total logistics cost, but also the speed of cash flow and the value of capital. This concern also represents a direction for future research.

Finally, although our optimization models are able to identify the most economic transportation mode, the final decision regarding transportation mode preference depends on numerous external and internal factors. For example, China's transportation infrastructure, additional risk factors, and Company P's relationship with freight forwarders within and outside of the U.S., to name just a few, can affect the final decision significantly.

5. Concluding Remarks

With outsourcing becoming an integral part of a corporation's pursuit of competitive advantages, a clear understanding of all hidden costs associated with this strategy has received a great deal of attention. Company P, one of the leaders in the aviation industry, formed a joint venture in Chengdu, China in 1996 and has been deeply concerned with the costs and benefits resulting from outsourcing some of the aircraft engine parts to this joint venture. This project focused on analyzing Company P's global logistics system and developing sound techniques that aid the companies in making effective transportation policy decisions. The major cost components of the logistics system that affect transportation policies were identified and serve as critical inputs in the decision-making process. A set of optimization models and solution methods for five potential transportation mode combinations were developed. The cost components, the results derived from the optimization models, and the use of sensitivity analysis on proposed solutions have provided helpful guidelines for Company P to evaluate the cost effectiveness of each potential transportation alternative.

Acknowledgments

The author would like to thank APICS-the E&R Foundation and Supply Chain Council for their financial support for this project (Grant #2000-3) and Mr. Christian Rossetti and Mr. Ken Mair for their assistance in collecting the data and information of the two companies involved in the project. The author would also like to extend her gratitude to the editors and two reviewers for their thoughtful and constructive comments, which have helped improve the quality of this chapter.

References

Anonymous (1999). A logistics checklist. *CMA Magazine,* page 20.

Ballou, R. (1999). *Business Logistics Management: Planning, Organizing, and Controlling the Supply Chain.* Prentice Hall, Inc., Upper Saddle River, New Jersey.

Bertazzi, L., Speranza, M., and Ukovich, W. (1995). Minimization of logistics costs with given frequencies. In *7th WCTR Proceedings, Vol. 4,* pages 45–56.

Fagan, M. (1991). A guide to global sourcing. *The Journal of Business Strategy,* pages 21–25.

Fera, P. (1998). Developing a framework for international outsourcing in the commercial aircraft industry. Master's thesis, Sloan School of Management, Massachusetts Institute of Technology, Cambridge, Massachusetts.

Lambert, D., Stock, J., and Ellram, L. (1998). *Fundamentals of Logistics Management.* The McGraw-Hill Companies, Inc.

Lee, C.-Y. (1986). The economic order quantity for freight discount costs. *IIE Transactions,* pages 318–320.

Liao, T. (1997). Modeling and cost analysis of global logistics and manufacturing system. Master's thesis, Department of Ocean Engineering, Massachusetts Institute of Technology, Cambridge, Massachusetts.

Maltz, A. and Ellram, L. (1997). Total cost of relationship: An analytical framework for the logistics outsourcing decision. *Journal of Business Logistics,* 18(1):45–66.

McGowan, K. (1997). The rise in global sourcing and logistics implications for shippers. *Transportation and Distribution,* pages 83–87.

Richardson, H. (1995). What value logistics. *Cost/Value Series, Part III, T&D.*

Russell, R. and Krajewski, L. J. (1991). Optimal purchase and transportation cost lot sizing for a single item. *Decision Sciences,* 22:940–952.

Silver, E., Pyke, D., and Peterson, R. (1998). *Inventory Management and Production Planning and Scheduling.* John Wiley & Sons, Inc.

Tersine, R. and Barman, S. (1991). Economic inventory/transport lot sizing with quantity and freight rate discounts. *Decision Sciences,* 22:1171–1179.

Tyworth, J. and Zeng, A. (1998). Estimating the effects of carrier transit-time performance on logistics cost and service. *Transportation Research (A),* 32(2):89–97.

van Damme, D. and van der Zon (1999). Activity based costing and decision support. *International Journal of Logistics Management,* 10(1): 71–82.

Vidal, C. and Goetschalckx, M. (1997). Strategic production-distribution models: A critical review with emphasis on global supply chain models. *European Journal of Operational Research,* 98:1–18.

Chapter 13

THE SUPPLY CHAIN IN THE FOREST INDUSTRY: MODELS AND LINKAGES

Andrés Weintraub
Department of Industrial Engineering
University of Chile
aweintraub@dii.uchile.cl

Rafael Epstein
Department of Industrial Engineering
University of Chile
repstein@dii.uchile.cl

Abstract We discuss the supply chain of the Chilean forestry sector. The chain starts with a basic raw material, consisting of standing trees in forests. The products, at different stages of transformation, then proceed to local markets or are exported: as logs, lumber, manufactured products, pulp and paper. Processes involve primary and secondary destinations for the logs, with transformations carried out at different plants, such as sawmills and pulp plants. Transportation plays an important role in moving products along the chain. A series of mathematical models have been successfully implemented along the chain to support decisions. We discuss the flow of material along the chain, as well as links between decision making processes, pointing out where improvements should be made to improve coordination.

Keywords: forestry, supply chain, mathematical models, logistics

1. Introduction

For forest products, the production chain starts with standing trees in the forest and then continues on to one or more stages of transformation, before goods are sent out to customers as final products. We will describe the forestry supply chain from the perspective of Chilean forestry firms,

J. Geunes et al. (eds.),
Supply Chain Management: Models, Applications, and Research Directions, 343—362.

whose final products go to local markets (pulp and paper, lumber or sawn pieces mostly) or are exported in the form of logs, lumber, plywood, manufactured pieces, paper, or pulp. The industry relies mainly on pine plantations with a growth cycle of about 25 years, located in southern Chile.

The shape of pine trees defines their best possible use. The lowest, widest part of the tree is the most valuable, and is either exported as logs or used by local sawmills for higher priced manufactured products. The tree's mid-section, of medium diameter, is sent to local sawmills for less valuable lumber. The highest part of the tree, which is narrowest and least valuable, is used for the pulp mill. Sometimes downgrading of timber occurs: that is, thicker logs may be used for lower-value products at a loss. Combining these end uses makes for the best use of forests, because the whole tree is used appropriately.

The supply chain starts with logging operations in the forest, where harvesting machinery brings logs of different sizes, already cut to certain dimensions, to the roadside. Trucks take loads of these logs to their first destination, which may be:

a. Pulp and paper mills. Firms have one or two pulp mills usually located by the coast for technological reasons, given the need for large quantities of water. Most paper mills use pulp to produce paper, but a few use logs to produce paper by a mechanical process.

b. Sawmills and plywood plants. A typical firm has several sawmills, each with different characteristics, spread out among forest plantations to minimize transportation costs. Plywood plants are also a first destination for logs.

c. Ports, where logs are sent directly for export. These logs disappear from our supply chain once they are loaded onto a ship.

d. Stockyards, where logs are kept over the short term, or from summer to the rainy winter season, then sent on to their next destination.

e. A collection-transformation center, located near pulp mills, where some of the logs that have not been cut to final size in the field are sent for cutting and sorting. From there, the logs are sent on to their next destination.

Products go from these initial destinations on to secondary destinations. Boards of different size require transportation to go from mills to local markets or ports for export.

Other boards go to manufacturing plants, usually located near mills, for manufacturing into more complex final products, to be used in building and industry. These manufactured pieces again go to local markets or to port for export.

From a pulp mill the product, pulp, goes to a paper plant for further manufacturing into paper, or is exported directly. Paper is mainly used for the local market but some is exported. Plywood is also oriented to international markets.

A set of systems based on mathematical models have been implemented and are being used at different levels of the supply chain to support operating decisions. We will describe how these models support decisions at this level, analyzing the linkages between each stage. Figure 13.1 provides a basic diagram of a supply chain.

Section 2 shows two systems developed and implemented at the forest level. One is a short-term harvesting model, to decide which areas to harvest, how to cut the trees, and where to send them to satisfy short-term demand. The second system decides the location of harvesting machinery and access roads. Both systems are used by many firms in Chile.

Section 3 describes a system based on a simulation model using heuristics, implemented to handle forest to primary destination transport, and transport from mills to final destinations. Again, this system is heavily used by forest firms.

Section 4 describes secondary processes, mainly related to sawmills and remanufacturing and includes the description of a scheduling model, which has not been implemented.

Section 5 discusses the overall supply chain, as shown in Figure 13.1, and the links between each stage, in terms of decisions regarding material, information and communication flows. In particular, we discuss the limitations in the supply chain in terms of coordinating different stages and the need to introduce technology, such as GPS, bar codes, and real time flow of information, to improve overall performance.

We note that in most cases there is mixed ownership of assets. Major firms own most of the timber lands harvested, but also purchase timber from small landowners. They own pulp, industrial, and paper plants and most of the sawmills, but third-party mill operations are also hired. Timber harvesting and transportation is mainly subcontracted.

2. The supply of timber

Recall that the basic raw material is standing timber. Large firms own from 50,000 to 400,000 hectares of pine plantations. At the operating

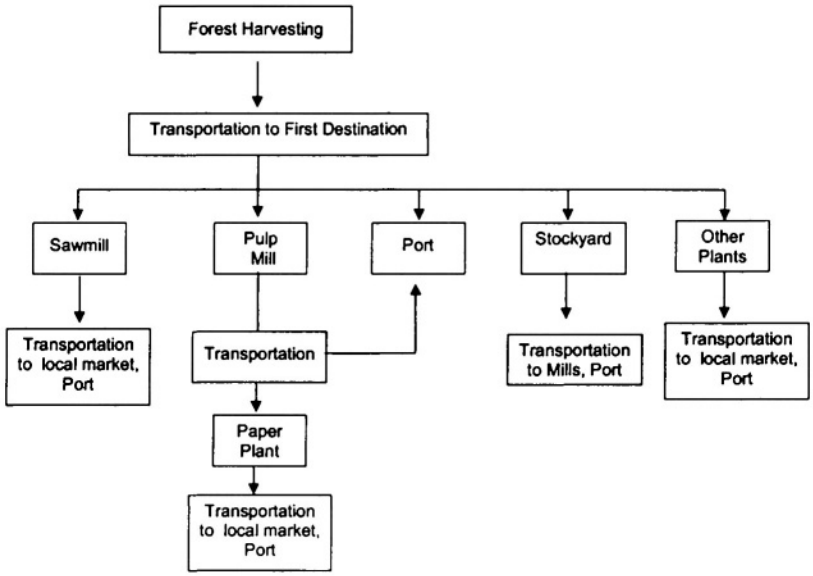


Figure 13.1. The Supply Chain

level, at any moment some areas have mature timber ready for harvesting and for which access roads have already been built.

For the purpose of harvesting, forests are divided into reasonably homogeneous areas, called stands, characterized by similar tree age, site quality and management conditions, which lead to similar tree lengths and diameter. On any given day, harvesting goes on in a number of stands or points within the forest. The harvesting process starts when loggers fell a tree. Once on the ground, the branches of the tree are cut off. Next the tree is cut into pieces according to instruction so as to obtain the products demanded by length and diameter (bucking) or the whole log can be sent to a sorting center for bucking.

The loggers receive instructions on how to buck trees in a given area. Figure 13.2 shows a simplified example, where a first cut is 12 m long and at least 27 cm in diameter, yielding an export log. A second cut is 4 m in length and at least 22 cm in diameter (sawmill log) and the rest of the tree, with a diameter of over 8 cm, goes to a pulp plant. Inventory simulators based on sample plots estimate the volume to be obtained for each product, given a bucking pattern.

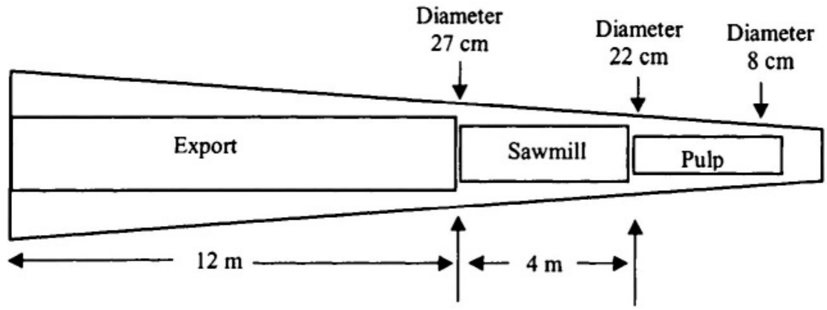


Figure 13.2. Simplified Example of Bucking

Basic decisions at this level are:

a. Which stands will be harvested in each period (a week is a typical period here).

b. The bucking or tree cutting instructions to be used for each stand and period.

c. The products to be sent to different destinations, to satisfy specific demands for length and diameter during each period.

d. General needs for harvesting machinery and trucks.

OPTICORT (Epstein et al. 1999), a Linear Programming (LP) model, was developed to support these decisions.

OPTICORT makes decisions on how to produce the many different products, as defined by the length and diameter needed each week to satisfy varying demands for exports and at sawmills and pulp plants. Loggers are given instructions on how to buck or cut every tree, to obtain up to seven pieces according to a list. These instructions are called bucking patterns and use information from an inventory simulator that indicates how many m^3/ha will be obtained for each defined product in a given stand using any given bucking pattern.

A basic, simplified structure of the model defines two variables.

Y_{idkt} = Volume of timber (m^3) transported from stand i to destination d, of product k in period t

K_{ijt} = Volume of timber (m^3) produced in stand i, with bucking pattern j in period t.

The objective is to maximize net returns:

$$\sum_{i=1} R_{idkt} Y_{idkt} - \sum_{ijt} Cost_i K_{ijt} \qquad (13.1)$$

where R_{idkt} represents sales revenues minus transportation costs and $Cost_i$ denotes harvesting costs.

The main constraints are:

$$\sum_{j,t} K_{ijt} \leq Vol_i \qquad \text{for all} \quad i \qquad (13.2)$$

so that total volume harvested is limited to existing timber Vol_i

$$\sum_{i,j} K_{ijt} \leq CAC_t \qquad \text{for all} \quad t, \qquad (13.3)$$

so that timber harvested is limited by machine capacity, CAC_t, and

$$\sum_{j} REN_{ijk} K_{ijt} - \sum_{d} Y_{idkt} \geq 0 \qquad \text{for all} \quad i, k, t, \qquad (13.4)$$

as volume transported is limited by production of each product by stand and period. The parameter REN_{ijk} is the fraction of volume for product k, when using bucking pattern j in stand i, which is obtained from an inventory simulator based on sampling techniques.

The problem is a standard LP, except for the extremely large number of possible bucking patterns. For this purpose, column generation was used to generate branching patterns. Given an optimal solution for a set of defined bucking patterns, additional bucking patterns are created through a specially developed branch and bound column generation subroutine. At the source node, the first branches consist of all possible first cuts by length and diameter (e.g 12m/26cm). These create a set of 10 to 30 branches (compared to two branches in a 0-1 branch and bound scheme). At most a tree can have seven depth levels, which constitute the maximum number of possible cuts in a bucking pattern. A path in the branch and bound tree corresponds to a bucking pattern.

The dual variables of the present LP and the yields per product from the simulator lead to the value of each node in the branching.

The model works with demand data provided in the form of inputs based on the estimated needs of sawmills, pulp mills and exports. Exports of logs are defined by contracts, while mill requirements are defined by estimates of requirements, which in turn are driven by estimates of demand for final products.

Data on harvesting and transportation cost are based on experience, as well as results provided by a machine-location model and a transportation system, discussed below.

Several Chilean firms use this model and a large Brazilian firm uses a similar model. The model runs in about ten minutes on a personal computer, using a commercial LP code, and has led to significant savings. The model's main advantage is that it identifies a good match between product demand and standing timber, which significantly reduces the downgrading of higher diameter logs and the cutting of excessive timber.

Based on model indications for the first week and feedback from operations, operations are scheduled daily. This involves scheduling daily production from each source, the exact daily volume for each product obtained through bucking, and how daily demand will be met. A weekly program is basically the combination of daily programs, which will vary from day to day, due to operating dynamics: schedules are not met exactly, external conditions such as demand may vary daily, or unforeseen events, such as changing weather conditions, may occur.

At present, daily operations are scheduled manually. A scheduling LP model to support daily decisions will be developed in the near future.

A second model to support harvesting deals with the following problem. Given an area of say, 500 hectares, scheduled to be harvested in the next several months, where should harvesting machinery be located and what secondary roads should be built? There are basically two types of harvesting machinery used to bring felled logs to the roadside. On flat terrain, skidders or tractors are used, while for steeper slopes cable logging or towers are used to bring logs up hill or, in some cases, downhill.

Short access roads connect these operations to existing primary roads. Given the slow speed of skidders, it is not considered economical to use them for distances of over 300 meters from roadside. This defines the road requirements for flat areas.

For steep areas, wherever a cable logging operation is installed, an access road must be built, along with a flat area to stock logs before loading them into trucks.

A decision making system, PLANEX (Epstein, et al. 1999) was developed to support these decisions. The system uses data provided by a GIS for information on timber existence, topographical height levels, and geographical features, such as rivers, and existing roads. The information given by the GIS, typically in vector format, is converted to a raster format with cells, typically of 10×10 meters. The user, normally the planning engineer, provides additional information on the productivity of harvesting machinery and road building, harvesting and transportation costs. Thus, the data is derived from users and GIS information. An interactive graphic interface helps the user observe and modify solutions. The determination of solutions is based on heuristics.

In a first step, a greedy approach selects best locations sequentially. For any given (and still available) location, the timber and revenues that can be obtained by a machine at this location are determined using GIS information and harvesting costs. GIS information on topography is used to evaluate possible roads to be built, while considering constraints such as maximum slope and turn radius. A shortest-path algorithm on this network builds the minimum cost path from the location to an existing road (transportation costs are negligible given the short distances involved). The location with best net benefits is chosen, cells harvested through it are excluded from further consideration, and the path thus defined is used as existing for further iterations. When enough locations are defined to extract all profitable timber, a local exchange algorithm is used to improve locations. Finally, the ultimate set of roads to be used is determined using a spanning tree that connects all chosen locations.

Several Chilean forest firms use this system and it is also being installed in Colombia at present. It has provided significant improvements over manual approaches using maps. The main result is the need to build fewer roads, which offers both monetary savings and better environmental impacts.

Some of the information provided by PLANEX on harvesting costs and productivity is an input to the short-term OPTICORT harvesting model.

Note also the interaction between the two models. The machine location system feeds cost and productivity information into the short-term harvesting system, while the latter supports decisions on which stands to harvest. We will discuss this point further in Section 5. Note also that the machine location model indicates positions for the machinery but not timing. The short-term harvesting model will indicate when to use skidders and towers in each respective stand, and includes limitations on the availability of each type of machine.

The machine location problem can be seen as a combination of an uncapacitated plant location problem, where the machines correspond to plants, and cells containing timber act as customers, combined with a network design problem when looking for the best road network. Several researchers have attempted to solve a mathematical formulation of this problem. Guignard et al. 1996 attempted to solve a formulation of such a model using Lagrangian Relaxation and strengthening the LP formulation. Epstein et al. (2000) applied a dual strategy to an expanded formulation of the problem. We expect better algorithms from this research in the near future.

3. Transportation to first destination

As described in Section 1, daily harvesting operations lead to logs being delivered at the roadside. In the case of skidders, these bring the logs to specified points along the road, where they are stacked and a loader machine loads them onto trucks. At tower locations, a loader installed in the flat area will do the loading. These logs are already sized by length and diameter and planners know their possible destinations: port, sawmill, pulp mill, industrial plant, paper plant, sorting center or stockyards. Usually, at each location stock remains from the previous day and production is scheduled throughout the day. At destinations, specific demand for products is organized according to priorities. Sending logs for export to Japan on a ship that is leaving in three days naturally has a higher priority than demand at a mill with safety stock. (We will discuss the issue of safety stock in Section 5.) Forest firms work with a fleet of from 50 to 300 subcontracted trucks depending on the size of operations. Loading or unloading takes about 20 minutes, while trips last from 30 minutes to several hours. Trucks travel between multiple origins in the forest and different destinations. A system, ASICAM (Weintraub et al. 1996), was developed to schedule daily trips. Truck scheduling is carried out at a transportation center typically located close to the forest firm management.

The system is based on a simulation model driven by heuristics. The model takes as inputs:

a. Demand for each product, at each destination, with priorities for next day.

b. Next day availability of stock of each product at each origin.

c. Scheduled production of each product at each origin for the next day.

d. Loading and unloading capacities at origin and destination.

e. The fleet of trucks available (not in maintenance) by type of truck, which determines load capacity and type of logs trucks can carry (e.g. regular trucks can only carry shorter logs) and operating conditions (some trucks cannot climb steep slopes when loaded).

f. Times and costs for loading at origin and unloading at destination, loaded travel from origins to destinations, and unloaded travel back to origins.

With this information the model simulates the operation. At 6.30 a.m. for example, loading starts at origins. Besides existing stocks, production for each product at origins is assumed to arrive uniformly throughout, for example from 9 a.m. to 4 p.m., as defined by the scheduler. After trucks are loaded, they leave for destinations and new trucks start loading. Once trucks reach their destinations, they receive a new assignment: go to an origin, load, travel to and then unload at a new destination. Trucks must queue if loaders are busy. Heuristic rules were derived for assignments. After extensive testing, a set of heuristics was found that proved robust and led to efficient solutions. Under heuristic rules, trucks that will become free in the next hour are scheduled jointly, to avoid shortsighted solutions. A 15-minute rolling horizon is used. Rules basically pick trips for trucks that satisfy urgent demand for product delivery and minimize traveling and queuing costs. They also avoid causing queuing congestion affecting other trucks. The simulation proceeds through the day, assigning trips to trucks. Its results indicate:

a. Needs for trucks and loaders the next day.

b. Specific destination of products (the same product can be produced at different origins).

c. A schedule for each truck and loader for the next day.

A mathematical model has also been developed to solve this problem. The problem is based on two submodels, one defining the trips to be made, the other assigning trucks to trips. The models interact through a Lagrangian decomposition approach (Equi et al. 1997). The current CPU requirements to solve this model, however, make it uncompetitive with the heuristic method already in use, which takes just a few minutes. Such a model approach could become useful in the future to integrate daily production and transportation.

Schedules are delivered by fax or e-mail to drivers and loader operators early in the afternoon. During the day, the transport center ensures that drivers keep to their schedules (a driver who does not arrive on time must wait). It also deals with deviations from the program. About 80% to 90% of the trips are carried out as scheduled. Unforeseen events, such as a truck or loader breaking down, cause deviations from the schedule, which at present are handled manually by the transport operator. Changes in schedule are communicated by radio.

Most Chilean forestry firms are using the system, with very significant savings of about 15% of total costs. Brazilian and South African firms also use it, and the latter won the South African Logistics Prize of 1996 using this system.

It has also helped the supply chain, as the system's use has compelled forest managers to keep a much tighter control on stocks in the forest, thus reducing losses and thefts. The system has improved companies' ability to satisfy demand for products at destinations, as well as producing a steadier stream of arrivals, which has helped to synchronize downstream operations. We believe the system has also helped forestry companies to significantly decrease their safety stocks at destinations.

The system now is in the first stage of making decisions in real time. This requires the use of GPS and real time communication between trucks, loaders, and the transport center providing data and instructions.

We note the interaction between the transportation system and harvesting decisions.

- By running the transportation system under different scenarios, the firm has information on the needs for loaders and trucks for the short term harvesting problem.

- The daily harvesting schedule and the transport schedule must be consistent. At present, harvesting schedules are carried out first and then results are sent to the transport center. The schedule allows flexibility by setting priorities. The transport center usually makes several runs to evaluate different transport and demand satisfaction scenarios and it can decide, for example, to add a few more trucks to carry additional timber products.

- The expected level of timber production will indicate the basic dimensions of the fleet. Based on this information, annual contracts are signed with trucks owners.

4. Secondary Processes

In this section we describe downstream operations. This description will include the necessary transportation. A description of these processes follows, with its linkages to the system, and how decisions are made at each level.

a. **Stockyards.** In a seasonal cycle, stocks are accumulated in the summer to cover the rainy winter season. Stocking up in summer allows using cheaper dirt roads to access forest areas. Part of the harvested timber is carried to stockyards, located nearer to secondary process facilities, with access by gravel roads, which can be used year-round. At the operating level, stockyards, in particular those in pulp plants and sawmills, act as safety stocks. Typical stocks are enough for a few days' use up to several weeks. On a daily basis, therefore, some products will be shipped to stockyards, while other will be shipped out. Forest firms also use stockyards as centers for buying timber.

b. **Sorting Yard.** A second intermediate type of process is the sorting yard, like that of Bosques Arauco, the largest Chilean forestry firm. Located in the pulp mill area, it receives whole trees (without bark, branches and top) which are bucked there automatically, according to the needs of different destinations, which are either in the area or close by. The advantage of the sorting yard over bucking in the forest is that logs, which in the forest would be subject to one bucking pattern, will be analyzed individually, and bucked according to the best use for that particular tree, given its dimensions, curvature, defects and needs at nearby destinations. In addition, automatic sorting yards provide products of better quality than those bucked manually. Transportation costs, however, rise due to the intermediate station thus introduced. Scheduling of additional transportation to next destination is done manually and specially assigned trucks are used.

c. **Sawmills.** Large forestry firms handle a set of sawmills. The sawmills usually receive logs that are four meters long, with diameters starting at 16 cm. There are different types of sawmills, which vary by level of technology and ability to handle different log diameters, and produce different products, but there is a basic process pattern. The logs go first to a primary cut, producing outer, more valuable boards. These boards are cut by rotating the log and cutting pieces, called lateral boards, to given specifications. The remainder of the log, called the central piece, can be sold as is, or go on to a secondary process, where it is cut into thin boards. Depending on the quality of the secondary sawing process, these lower quality boards can have different levels of smoothness. Again, the secondary board can be sold as is, or go on to drying and in some cases sanding before being cut into smaller boards of different sizes, for example 20 cm × 4 cm × 0.8 cm, before being sent to market. The higher quality lateral boards may go to remanufacturing plants, where they are further processed into the more elaborate pieces used in construction or furniture. In both sawing processes a significant percentage of the log turns into chips and sawdust, which are used for fuel at sawmills and pulp mills.

Typically the sawmilling operation has a reliable demand forecast for final products, over a six week horizon. There are hundreds of final products, defined by length, width and diameter, as well as remanufactured products. Sales and production departments communicate directly, so that orders can be assigned to specific sawmills, and then go on to the drying and sanding processes.

In contrast, coordination of timber inputs is ad hoc. Sawmills take log input for granted and forests plan the type and volume of logs to send to each sawmill based on general notions, such as, "send thicker logs to such-and-such" a sawmill. Short-term harvesting and distribution need to become more integrated into sawmill planning and production.

A linear programming model was developed to optimize the assignment of specific tasks for each sawmill and other processes over a six-week horizon, given orders (local market and exports) and log inflow to mills. A prototype test was carried out, where data for expected sales and log stocks at sawmills was input manually.

Results of the test were successful, showing a better use of logs and satisfaction of orders, as the model fully exploits the characteristics and relative advantages of each sawmill. However, the model has not yet been implemented, as the firm is developing a data system to feed on-line information on sales and log movements directly into the model. Given the high quantity of data involved, trying to feed data into the model manually in a real daily to weekly planning situation is unrealistic.

The model basically follows the flow of logs as they are processed by the sawmill, as shown in Figure 13.3 with flow conservation and transformation. Decision variables start with the selection of logs to be used, from among those available in stock. In the primary sawmill process, a simulation model embedded as a subroutine determines probabilistically the expected number of specific pieces to be obtained for each cutting pattern. The user specifies a certain number of possible cutting patterns and the model selects specific cutting patterns to be used. In a similar way, decisions in the model are formulated for secondary processes, drying, sanding, use of third-party processes and purchases, and the transportation of products.

The resulting LP model is a standard one, with a large number of constraints and variables, due to the many different processes involved. Different options for logs sent to sawmills can be analyzed by running the model with several log stocking options for each mill.

The model is designed to run on a weekly rolling-horizon basis. Each week, information is added on new orders and past production. While adding new orders to sawmills (within their capacities) causes no problem, as a rule taking away orders not yet produced is not considered an acceptable practice, so the model inputs exist-

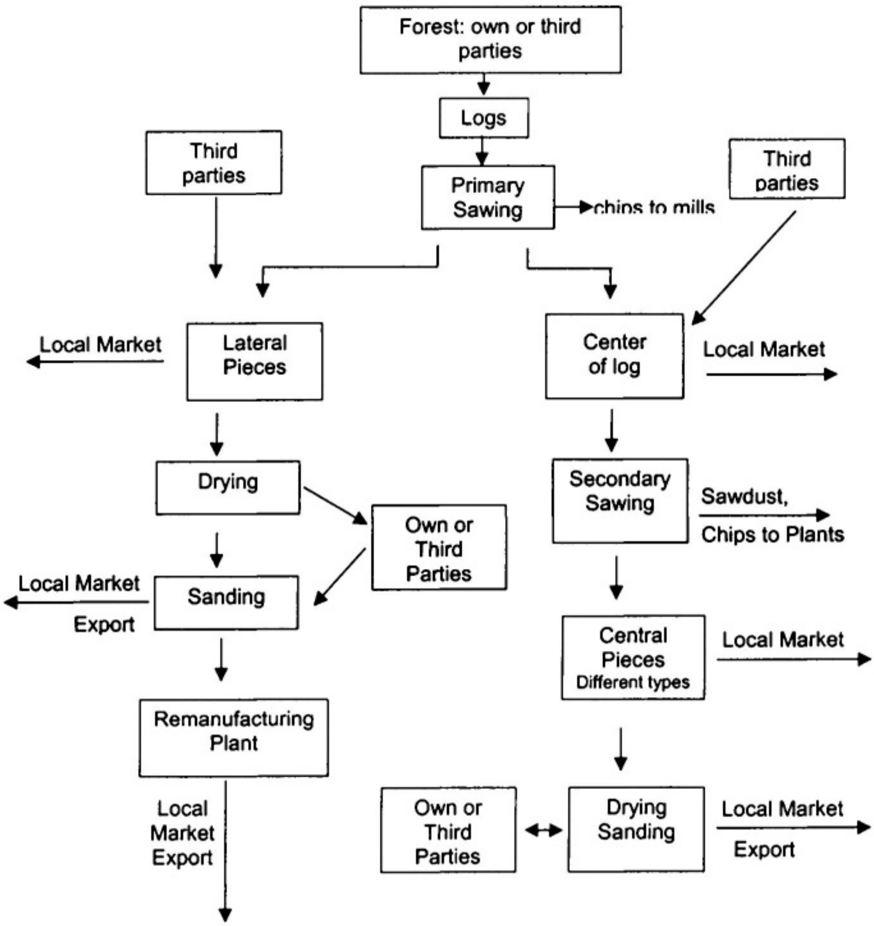


Figure 13.3. Diagram of the Sawmilling Process

ing assignments from previous runs and imposes a heavy penalty on such modifications. Some parameters are difficult to determine, e.g., the time required for different sawmilling operations, which in strict rigor depends on the surface and length of boards.

Since these parameters were not well-known, we used estimates based on average performance. The model used only user-selected cutting patterns. A column generation process, similar to the one used in OPTICORT, could be added in the future.

The largest sawmill company in Chile schedules transportation of end products to clients using a modified version of ASICAM. They report significant savings, as well as better performance, by using computerized decision models to coordinate transportation. The company decided to incorporate this tool after its successful experience with forestry operations.

d. **Remanufacturing plants** are usually located near sawmills. They receive as inputs high quality pieces from lateral stock, in given sizes, e.g., 20 cm × 6 cm × 0.8 cm, and process them into specific products used in carpentry, housing, furniture and industry. There is a specific demand for these products, which are both exported to meet specific contracts and sent to local markets.

e. **Other plants**, of secondary value in monetary terms, take in timber products and by-products like chips and sawdust to produce products such as fiber board for local markets and export.

f. **Pulp plants** are the core of the operation, financially. Logs are fed into the mill, which needs constant input. As mentioned, pulp mills take in thinner, less valuable logs normally, but if needed to satisfy mill demand, thicker logs may also be used, at a loss. Pulp plants are located on the coast, as they need a lot of water and waste disposal possibilities. Presently, reprocessing within plants allows for relatively clean residues. Moreover, being near ports reduces transportation costs, as most of products are exported. Pulp mills produce pulp, which can be exported directly, or sent to a local plant to be transformed into paper. There are different qualities of paper: newsprint, wrapping, etc, which go through different processes and plants before being sent to market.

g. **Paper Plants** use pulp to produce paper by chemical processes. Other plants use a mechanical process to produce paper directly from logs, then sell it on local and international markets. Typically newsprint is produced this way.

h. Plywood Plants are increasing in the Chilean forestry sector. New plants produce high quality plywood oriented for international markets.

5. Description of the Supply Chain

We have described the supply chain for forestry firms with its basic elements: harvesting, plants, sawmills, stockyards and the exit from the supply chain through ports for export or local markets for the different products in the chain: logs, lumber, manufactured pieces, other products, pulp and paper.

At most stages, mathematical models play an important role in decision making. It has been shown that the use of systems based on operations research models can bring significant improvement to each of the processes described. These systems won the 1998 Edelman Price of INFORMS, the US Society for Management Science and Operations Research (Epstein et al. 1999).

There are significant differences with typical supply chains, as described in Simchi-Levi et al. 2000, for example. One difference is ownership. Third parties along the chain exist (owners of timber land, mills, kilns, etc.) but play a relatively minor role. The main processes in the chain, i.e., timberlands, sawmills, and pulp mills, belong legally to different firms, and are operated by independent management; but, since they belong to the same holding company, management coordinates the different firms. Of particular interest is setting transfer prices along the chain; these are decided at the corporate level. It should be noted that the high cost of log transportation creates certain monopsonies in timber purchasing. As we will discuss, some links between supply chain components are weak, particularly the transmission of information and coordination of decisions. Figure 13.4 summarizes the physical supply chain.

The arcs joining processes correspond to transportation activities. Stocks are held at most stages. In the forest, standing timber in mature and accessible areas can be treated as inventory, which can be obtained by harvesting at any time. The policy of a Brazilian firm, Aracruz, is to leave a minimum amount of standing timber in easily accessible areas as safety stock. There are also small stocks of felled trees in the forest, usually to start daily operations.

Stockyards store timber, mostly from summer to winter. Stocks in the form of logs of the necessary size are held at the different sawmills and pulp plants before processing. These safety stocks, which cover needs from several days to weeks, are not used to cover demand uncertainty,

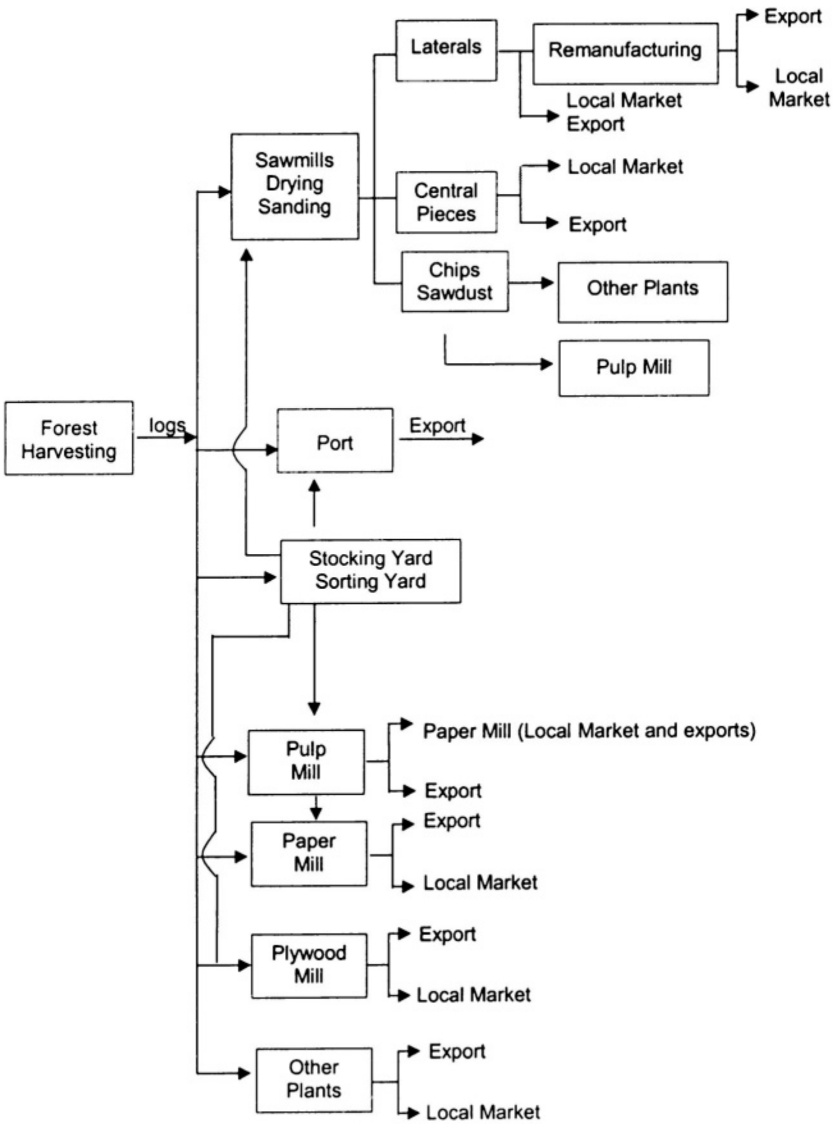


Figure 13.4. The Physical Forest Supply Chain

but rather to cover for poor weather or other contingencies that may hold up log supplies. Another reason for the stocks is the need to cover for weaknesses in coordination.

Stocks of intermediate products are held in the sawmilling process, before drying and sanding for example. Finally, stocks of final products, from logs to manufactured products, are also kept, before shipping. For example if we consider a ship to be loaded with 20,000 m^3 of different products: given the speed of harvesting and transportation, stocking of products at port should start ten days before the ship arrives.

In this case, stocks and production along the supply chain are driven by demand information, which is relatively reliable. This means the Bullwhip Effect (Lee et al. 1997) is not an important factor. Demand uncertainties usually arise from spot sales, which are transmitted immediately to the plants. For their part, forest operations are flexible enough to react to modified orders when needed. Moreover, downgrading, the practice of using thicker logs when thinner logs are unavailable, also adds flexibility to plant supply (Gavirneni et al. 1999).

It is important to consider the level of information and coordination quality in the supply chain. Following the description in Figure 13.1 we can point out the weaker points in the chain:

a. Improvements in the daily coordination of timber harvesting and transportation to first destination would strengthen the integration of both activities. In this form, jointly planning daily production and transportation would improve the overall process.

b. There is little coordination between decisions on harvesting to supply logs to sawmills and the processing required to satisfy known demand for sawn products. A proposed model that integrates these two aspects is presented in Lidén and Rönnqvist (2000). It is important to integrate the decisions of log supply to sawmills with the demand for end products and mill processes, to maximize overall revenue, including timber value.

c. Sales demand and sawmill planning are poorly integrated, given the rudimentary, manual planning process presently used. A model to better coordinate sales and sawmill production was described above.

d. Transport decisions, aside from their lack of integration in harvesting, are not carried out in real time, leading to loss of flexibility. The main advantage of real time scheduling would be "just in time" dispatching from the forest, as timber availability, communicated in real time, makes it possible to reschedule trucks and harvesting during the day in response to new requirements. Technology in data transmission allows these improvements, and some preliminary efforts in this direction have been carried out in several countries with promising results. Rönnqvist and Ryan (1995) report on one case in New Zealand where a real time decision model to schedule trucks was used for a short period, but later abandoned.

6. Conclusion

We have described the physical supply chain in the case of Chilean forestry firms and discussed the use of models and computer systems at different stages, as well as the weak coordination between some stages.

If we look at a desired future system, with proper technology the whole supply chain could be linked in real time, integrating daily forest harvesting and transportation, short-term sawmill processes, scheduled sales, and log supply.

Bar codes on logs being loaded into trucks, containing specifications (length, diameter) and transmitting data instantaneously to a central scheduling office would keep track of logs throughout the process, permitting integrated decisions on harvesting, tree bucking, and transportation in real time. On a daily basis, it would also make sawmill handling more integrated and flexible. Given how the price of the necessary technology (GPS, GIS, and data communication) is falling, at least some of these technologies will likely be applied in the near future.

References

Epstein, R., N. Rodriguez and M. Olivares (2000), "A Dual Ascent Algorithm to Location Problem and Road Design in Forest Harvesting," University of Chile, Working Paper.

Epstein, R., R. Morales, J. Serón and A. Weintraub (1999),"Use of OR Systems in the Chilean Forestry Industry," *Interfaces* 29(1), 7-29.

Equi, L., G. Gallo, S. Marziale and A. Weintraub (1997), "A Combined Transportation on Scheduling Problem," *European Journal of Operational Research* 97, 94-104.

Gavirneni, Srinagesh, R. Kapuscinski and S. Tayur (1999), "Value of Information in Capacitated Supply Chains," *Management Science* 45(1), 16-24.

Guignard, M., G. Bravo, J. Vera and A. Weintraub (1996), "A Lagrangean Relaxation Approach to Location Problem in Forest Harvesting," VIII CLAIO, Annals, 531-536, Rio de Janeiro, Brazil, August.

Lee, H., V. Padmanabhan, and S. Whang (1997), "Information Distortion in a Supply Chain: The Bullwhip Effect," *Management Science* 43(4), 546-558.

Lidén, B., M. Rönnqvist (2000) "CustOpT - a Model for Customer Optimized Timber in the Wood Chain," LiTH-MAT-R-2000-15, Linköping University, Sweden, to appear in the Proceedings of the 12th NOFOMA Conference, June 14-15, Aarhus, Denmark.

Rönnqvist, R and D. Ryan (1995), "Solving Truck Despatch Problems in Real Time," Proceedings of the 31st Annual Conference of the Operational Research Society of New Zealand, August 31-September 1, Wellington, New Zealand, 165-172.

Simchi-Levi, D., P. Kaminsky and E. Simchi-Levi (2000), *Designing and Managing the Supply Chain,* Irwin, McGraw-Hill.

Weintraub, A., R. Epstein, R. Morales, J. Serón and P. Traverso (1996) "A Truck Scheduling System Improves Efficiency in the Forestry Industry," *Interfaces* 26(4), 1-12.

Chapter 14

THE BENEFITS OF INFORMATION SHARING IN A SUPPLY CHAIN: AN EXPLORATORY SIMULATION STUDY

Tonya Boone
PO Box 8795
School of Business
College of William and Mary
Williamsburg, VA 23185
Tonya.Boone@business.wm.edu

Ram Ganeshan
PO Box 8795
School of Business
College of William and Mary
Williamsburg, VA 23185
Ram.Ganeshan@business.wm.edu

Alan J. Stenger
Smeal College of Business Administration
Penn State University
509 Business Administration Building I
University Park, PA 16802
ajs@psu.edu

1. Introduction

A supply chain is a network of firms, activities, organizations, and technologies that performs the functions of procurement of material from vendor firms, transformation of this material into intermediate and finished products and the distribution of these finished products to customers. It is often easy to identify a supply chain in a manufacturing

J. Geunes et al. (eds.),
Supply Chain Management: Models, Applications, and Research Directions, 363–381.

enterprise, although the complexity of the chain may vary from industry to industry or even company to company. As an example, a wine producer has several sets of suppliers, with its vineyards, bottlers, cork and carton manufacturers and so on. These firms ship to the winery. Once the wine has been made and aged, the winery supplies a set of customers, consisting of retailers, pubs, hotels, wine-bars, as well as the ultimate consumers ordering via the Internet. In the book publishing industry, Barnes and Noble uses more than 20,000 suppliers, consisting of major publishers to small University presses, to stock over one million books at their Jamesburg, NJ distribution center. The distribution center replenishes retail locations daily based on point of sale data transmitted by stores. On the other hand, Barnes and Nobles Internet fulfillment center (bn.com) stocks and replenishes over one million books for immediate delivery to the customer.

As the above examples indicate, a supply chain consists of a number of entities interacting with each other in complex ways. A supply chain could have vendors; International Purchasing Organizations (IPOs) to procure raw materials; a variety of transportation options to ship them; numerous ways to produce the product; and finally several channels to distribute the product. Any firm can do some or all of these operations in-house or decide to outsource them, typically to Third Party Logistics (3PL) providers. One can also imagine the innumerable number of flows within and across these entities–the flow of product from the supplier organizations to the point of sale; the flow of information between supply chain entities such as orders, tracking requests, etc.; there are cash flows including invoice preparation and transacting payments; there are process and work flows that manage operations between these entities; and finally intra-firm collaborative teams constitute the people flows in the supply chain.

Research in managing many of the flows described–popularly referred to as supply chain management (SCM)–has grown exponentially over the last decade. If one culls out the various research streams in SCM, they align themselves in one of five broad categories: (i) competitive strategy, (ii) costs and benefits of information sharing, (iii) managing product variety (via product postponement), (iv) supply contracts, and (v) the economics and logistics of network location and optimization (for a comprehensive review of the literature, see Ganeshan et al., 1998). Our emphasis in this research is to study the importance of information sharing in a complex supply chain. Specifically, as ensuing sections will describe, our supply chain consists of four echelons with many entities in each echelon: the supplier, the plant, the distribution centers, and retail outlets. Within this framework we study, through a Simulation model, how collaborative planning, be it for forecasts, production plans, or replenishment between pre-determined entities of the supply chain produces value to the firm, the customer, and ultimately to the shareholder.

The idea behind supply chain management is to view the chain as a total system, and to fine-tune the decisions about how to operate the various components (firms, functions, technologies, and activities) in ways that produce the most desirable overall system performance in the long run. Doing so is extremely difficult due to the number and complexity of the decisions to be made, as well as the inter- and intra-organizational issues that must be addressed. Herein lies the dilemma for today's researchers. Should one model the complexity of the realistic supply chain? Doing so will, in all certainty, make the problem at hand intractable. On the other hand, one can simplify the models to get some key insights but run the risk of diluting the richness of the model to such an extent that it cannot be extrapolated to "realistic" scenarios. Simulation often provides the right middle ground to analyze such complex models. Although simulating the supply chain we are about to describe in this paper is quite a difficult task, it nevertheless provides us with a tool to analyze the impact of relevant parameters on supply chain performance.

The bulk of SCM research on the costs and benefits of information sharing, at least in the operations management realm, studies it from a myopic perspective. Specifically, since modeling the entire gamut of entities in the supply chain – i.e., from the suppliers to the end customers – is intractable, researchers often resort to stylistic models to study the costs and benefits of information sharing. The supply chain structure typically consists of arborescent structures typically limited to two echelons, a system of order transmission between these echelons (such as the reorder point system), and a simple, often inventory-related cost structure that encompasses the two echelons. The results and insights are often studied within these stylized environments (for a good overview on the modeling approaches see Tayur, Ganeshan, and Magazine, 1998, pages 337-465). Although such models are very practical, and are effective in providing insight into inventory-related supply chain performance, their primary shortcoming arises from the fact that their results, barring a few exceptions, cannot be easily extrapolated to realistic supply chains. Our objective in this chapter is to evaluate the value of information sharing, especially through simulating emerging supply chain initiatives such as Collaborative Planning, Forecasting, and Replenishment (CPFR), in a realistic supply chain. We base our analysis and results on data collected from a Fortune-500 company. This chapter, in addition to simulating

CPFR in realistic situations using realistic data, is unique in two other ways: one, we use the three key dimensions of supply chain performance oft-cited in the literature but seldom used together–customer service, time, and shareholder wealth. Second, as the methodology section will illustrate, our simulation includes most of the relevant costs and constraints, and captures the essential elements of product, information, and cash flows in a typical fast-moving consumer-goods supply chain.

The remainder of this chapter is organized the following way. In section 2, we will present the research hypothesis and the relevant performance measures we will be using. Section 3 describes our data, Section 4 the supply chain simulation and the methodology, and finally in Section 5 we provide a discussion and conclusions.

2. Research Hypothesis and Performance Measures

Research on information sharing in the supply chain was initiated by Forrester (1961) who demonstrated that information in a supply chain, such as orders, propagates upstream with increased volatility. Recently Lee, Padmanabhan, and Whang (1997a & b) have christened this phenomenon the "Bullwhip" effect. The bullwhip effect has the negative impact of increased inventory levels or large stock-outs for SKUs whose demands are volatile at the customer level. In an effort to curb the bullwhip effect, and to improve working capital efficiency, several firms have initiated programs that work towards sharing forecast and other planning information (for a discussion see Lee, Padmanabhan, and Whang, 1997a). The premise, of course, is that centralizing demand information will make all plans in the supply chain react to the same data, mitigating the bullwhip effect (Chen et al., 1998) and improving working capital efficiency. One example of such an information-sharing initiative is Vendor Managed Inventory (VMI). Under this initiative, the supplier or vendor is empowered to monitor and eventually replenish the customer's inventory according to pre-determined contractual agreements. Specific company examples include Barilla SpA (see HBS Case: 9-694-046) where inventory levels substantially reduced while maintaining high item fill-rates. Clark and Hammond (1997) show that the use of VMI in Campbell Soup has provided better performance gains. An example of vendor managed inventories in the grocery and consumer products industry commonly referred to as Continuous Replenishment Programs (CRP) is Pillsbury. The frozen foods division of Pillsbury maintains and monitors product and inventory flows to downstream retailers, such as Kroger and HE Butt (Long, 1999), enhancing inventory performance.

Efficient Customer Response (ECR) is another initiative to reduce volatility and uncertainty, primarily in the grocery industry. A key idea in ECR, in addition to reengineering the order management processes, involves sharing point-of-sale data between various links in the supply chain, enabling better replenishment, assortment planning, product introductions and promotions. Sharp and Hill (1998) estimate that ECR could potentially save more than 6% of sales in logistics costs and around 41% reduction in inventories for the grocery industry. Several other industries have adapted ECR to fit their unique needs: quick response (QR) in the apparel industry; efficient foodservice response (EFSR) in the foodservice industry; and efficient health consumer response (EHCR) in the medical/hospital supplies industry. All these initiatives have a common theme–share real-time information to improve working capital efficiencies and speed the product to the customer.

All the initiatives described above require some sort of technology to exchange information, which is typically done through Electronic Data Interchange (EDI) protocols. However, with the advent of the WWW, collaborative planning of forecasts and inventory replenishments (popularly called CPFR) can be done via the Internet. Initiated by the Voluntary Inter-industry Commerce Standards (VICS) association, CPFR promises a new business model the central theme of which is for businesses to align processes and standardize technologies to share forecast and other planning information securely, simultaneously, globally, and in real-time (see for example White, 1999). The key idea behind the CPFR initiative is that trading partners, e.g., the retailer and the manufacturer, collaboratively create forecasts (see Figure 14.1). Both the retailer and the manufacturer collect market intelligence on products; the retailer provides information on marketing programs, etc., and shares it in real-time over the Internet. In most cases, the retailer owns the sales forecast. If the manufacturer agrees with the forecast, automatic replenishments are made to the retailer via predetermined business contracts in order to maintain a specified inventory or customer service level. If the manufacturer and retailer cannot agree on the forecasts or if there are exceptions, such as an unusual demand season or a store opening, the forecasts are reconciled manually. The trading partners must agree on several key issues prior to implementing CPFR, such as how to measure service levels and stock-outs, how to set inventory and service targets. As the relationship develops over time, the retailer and manufacturer may jointly redesign key business processes in order to improve system performance. Several pilots of the CPFR business model are underway that allow retailers (e.g., Wal-Mart) and vendors (e.g., Lucent and Sara Lee) to "share information regarding key planning parameters (i.e., promotions, store

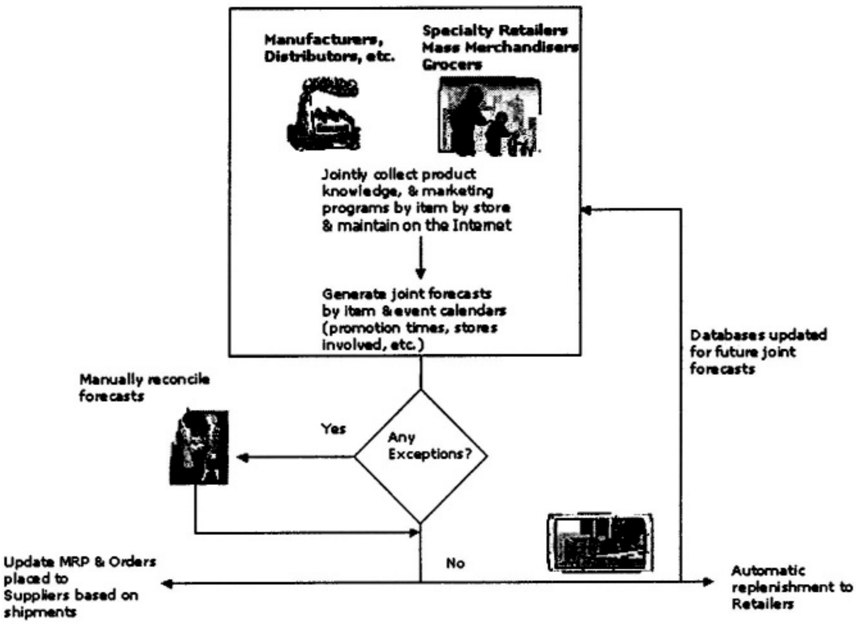


Figure 14.1. The CPFR Process

openings etc.) impacting forecasts; and communicate/resolve variances within item level forecasts." (from www.cpfr.org).

In contrast, in the ROP procedure, retail level planners collect product information and marketing programs at the store level (see Figure 14.2). Combining that with the point-of-sale (POS) data, item-level forecasts and event calendars that record promotion dates, special marketing programs, etc., are generated. Based on inventory and/or service level targets, the forecasts and the corresponding errors are used to generate reorder points. When inventory of a specified item reaches the reorder point, the retailer places an order to the manufacturer. If the product is available, it is shipped to the retailer; if not, the retailer will look for alternative solutions to replenish the item. The manufacturer, on the other hand, collects product knowledge and marketing programs of major retailers from public sources. Based on retailer orders and historical shipment information, the manufacturer generates a forecast by item, and in most cases, by retailer. These forecasts also drive the production of the items.

The research literature has little, if any, systematic process to determine the benefits of the CPFR process. Our intention is to simulate the process and compare it to traditional ordering and planning mech-

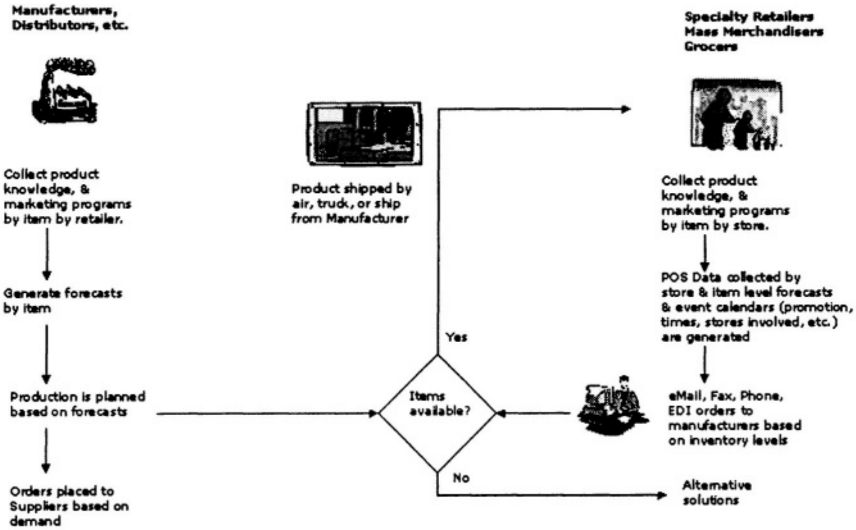


Figure 14.2. The "Traditional" Reorder Point Process (ROP)

anisms, specifically the reorder point system (ROP). We will measure performance on three dimensions - customer service levels (fill-rates), Shareholder Value (as Economic Value Added (EVA)), and the time dimension (as Supply Chain Cycle Time).

2.1 Performance Measures

Fill-rate: The firm in question, as the Data section will show, has six distribution centers (DCs) supplying to sixty-three retail markets. An overall fill-rate is computed using the volume-weighted average of the fill-rates at each of the DCs (for a similar measure, see Deuermeyer and Schwarz, 1981).

Economic Value Added (EVA): Since we do not model market forces and stock prices, we use the EVA, a measure developed by Stern Stewart as a measure of shareholder wealth. This is simply the profit less the true cost of capital. The profit is just the revenue less all the relevant costs that are involved in operating the supply chain. Capital is all the investment outlays incurred–including all infrastructure and technologies used in the supply chain.

Supply chain cycle time: This is defined as the total time spent by a product in the supply chain. It includes time at the supplier warehouse, transit to the plant, as raw material, WIP, and finished goods in the plant, transit to the DC, as inventory in the DC, and finally in transit before it reaches the retailer establishment. Since there is more than one entity in each echelon of the supply chain, time is just the volume-weighted average of the all the relevant times in a particular echelon.

2.2 Hypothesis

Our intention is to see if an information-sharing initiative such as CPFR has any tangible benefit for the above performance measures. To do this we will compare it to a traditional "reorder point" (ROP) methodology. Under the ROP method, the downstream facility in the supply chain (such as a DC) will place an order to its trading partner (such as a manufacturing plant) when reorder points, set independently, for different products are triggered. On the other hand, under the CPFR initiative, the trading partners (the DC and the manufacturing facility) will plan the replenishment quantities together, i.e., share information about demands. While benefits of such information sharing initiatives are proven in stylistic models, they have not been confirmed in a realistic and complex supply chain as the one we are about to describe. We would expect, even in the complex system, for CPFR to provide a higher level of fill-rate than the traditional reorder procedures. Formally,

H1: CPFR produces a higher fill-rate than the ROP method.

Since sharing of information produces better forecasts, the use of an information-sharing initiative reduces inventory in the supply chain. The hypothesis can be summed up as:

H2: CPFR results in lower supply chain inventories than the ROP method.

With accurate forecasts and low inventory, one can expect the supply chain cycle time or "response time" of the supply chain to be lower with the CPFR system:

H3: CPFR results in a lower supply chain cycle time when compared to the ROP method.

Finally, information-sharing initiatives are sustainable only if they add intrinsic value to the company and consequently to the shareholder. To the best of our knowledge, we have not seen the impact of information sharing on shareholder wealth, even in simplistic models. Since the use of CPFR-like initiative reduces working capital by reducing inventories, one would expect such initiatives to ultimately add shareholder value. Since we measure shareholder wealth by EVA, we frame the hypothesis as:

H4: CPFR results in higher EVA when compared to the ROP method.

3. The Data

The data used to fuel the simulation model is adapted from the supply chain of a Fortune-500 consumer products company (see Table 14.1). To maintain confidentiality, we have masked the data. Specifically, we have changed the nomenclature of the supply chain structure, i.e., the names of the markets, DCs, plant, and the suppliers are not the same as the original company. However the relative magnitude of the data is preserved. We chose one product from a product family of household cleaners that typically sells for around $2/pound.

The product in question can potentially be sold at retail locations in any one of sixty-three markets in the continental United States. These retail outlets are replenished by DCs via Less-Than-Truckload (LTL) shipments on a regular basis. Depending on the market location, and the supplying DC, the order cycle times (i.e., time from the retail order to the point of fulfillment by the DC) to these retail outlets range from one to five days. The freight rates from the DCs to the markets range from 0.0466 to 0.16 $/pound shipped (the data tables are quite cumbersome–for the sake of brevity, we have chosen not to show it). We divide the year into thirteen accounting periods, and each of these periods has twenty operating days. The demand is seasonal, peaking during the Spring-cleaning season.

Because of the large number of shipments to customers and the long distances to be covered, this company has several distribution centers (DCs). These are located in Los Angeles, California; Denver, Colorado; Dallas, Texas; Chicago, Illinois; Atlanta, Georgia; and Kansas City, Missouri. The cost structure is therefore piecewise linear, changing with the volume of product that is shipped through the DCs. The DCs are replenished through one of four modes of transport: Less-Than-Truckload (LTL), Truck Load (TL), Trailer or Container on Flat Car (TOFC/COFC), and Railbox Car shipments, each having a shipping capacity of 20,000; 40,000; 50,000; and 90,000 respectively.

The product in question is produced in one manufacturing facility, located in Denver, CO. For the purposes of this study, we assume that

Master Database	Sales Database	Performance Database
Products Price, Weight, Cube, Bill of Materials **Supply Chain Infrastructure** Names & locations of retail markets, DCs, warehouses, manufacturing facilities, suppliers. Network Design Data (who ships to who) The investment, fixed, & variable costs of operating the infrastructure for varying throughput levels. **Transportation Data** Transport modes available, their capacity, and lead time (mean and standard deviation) for every link in the supply chain Freight costs for every feasible link in the supply chain. **Other Cost data** Inventory holding costs Administrative costs (fixed variable, overhead)	Total yearly sales Retail market size (proportion of total sal Seasonality by period Forecast errors **Status Database** **Open Orders** Distribution Centers, Plant Warehouse, manufacturing facility **Inventories** Finished Goods (at DCs, warehouse) WIP (at plant) Raw Materials (plant in-bound) **Tracking Sales Vs. Forecasts** **Current Plans** Distribution Requirements at DCs Shipment plans at plant warehouse & suppliers Production plan at the plant	**Customer Service** Fill rates for every DC Time through the supply chain Total Sales **Inventories** Minimum, Maximum, average levels at each facility **Transportation** Lead time statistics for every mode **Costs** Costs to operate each facility in the supply chain Inventory costs at each facility Freight costs at every link Administrative costs **Financial Measures** Profits EVA

Table 14.1. The Data

the Denver manufacturing facility has enough capacity to satisfy the downstream demand. The initial investment to build and get the plant running was $13 million. Once the product is produced, it is stored in an "out-bound" plant warehouse (different from the DC) adjoining the facility. The operating economics of the plant-warehouse are, as before, piecewise linear and depend on the throughput. The product requires three raw materials, Cans/Bottles, Corrugated, and Chemicals, that make up 10, 30, and 60 percent of the product. Each of these raw materials is sourced from three major suppliers, located along the Gulf Coast and the Mississippi river. Each of these suppliers charges a different price, based largely on the quantity that is ordered and the delivery performance that is promised. The shipments from each of these suppliers to the Denver plant can be made through four available modes of transport–LTL, TL, COFC/TOFC, and Rail Boxcar. The freight rates, mean and standard deviation of transit times for these modes are also available to us.

Describing our data requirements in such detail, we believe, is valuable in two respects: one, it provides the reader with a list of data requirements that will be needed to simulate a realistic supply chain. Of course such data requirements will not be homogenous from company to company, but we believe that most retail distribution channels for fast moving consumer goods are likely to have similar data needs. Second, the reader can perhaps appreciate the difficulty in obtaining the data to simulate a supply chain. Even the addition of new transport modes to the analysis will exponentially increase the data needs!

4. The Simulation

Simulating a complex chain is obviously a difficult endeavor. To accurately capture all the costs and constraints and to appropriately model the CPFR business model, we wrote our own simulation routines in Visual Basic, a programming platform for the Windows operating system. Figure 14.3 shows the general outline of the simulation. For the purposes of brevity and to enhance clarity, we keep our explanation of the details of the simulation to a minimum. The simulation first reads all the data required for a run. This includes the products, markets, and sales data; detailed data on the operating economics of each facility in the supply chain; the freight options, cost structure, and delivery performance of each of the transportation modes. We then input the structure of the supply chain, i.e., market, DC, plant, and supplier locations, and the corresponding arcs that connect these "nodes" in the supply chain. The underlying data and the supply chain structure will remain constant across different simulation runs.

Our planning horizon is thirteen periods (one year), each period consisting of twenty days. The simulation begins by forecasting the demand for each period at each of the DCs. The total average demand for each market is known, and therefore the annual average demand on each supplying DC is also available. Based on seasonality of each period, the average demand for each period in each DC can also be computed in a straightforward way. We use the following procedure to "simulate" a forecast in each of the periods at the distribution centers (see also Sridharan and Berry, 1990)

Forecast in period t = Sales in period t + forecast error,

where the forecast error is assumed to be a normal variable with mean zero and a variance, σ^2, that is estimated from company data as (see Stenger, 1994) aS_i^b, where S_i is the average sales in period i, and a and b are positive constants.

Under a traditional reorder point system (ROP), the planner at the DC will use the forecasted demand, forecast error, on-hand inventory, scheduled receipts, transport mode characteristics (lead-time performance and lot-size) and a pre-determined fill-rate target to compute reorder points via well-known inventory methods (see Silver and Peterson, 1987,

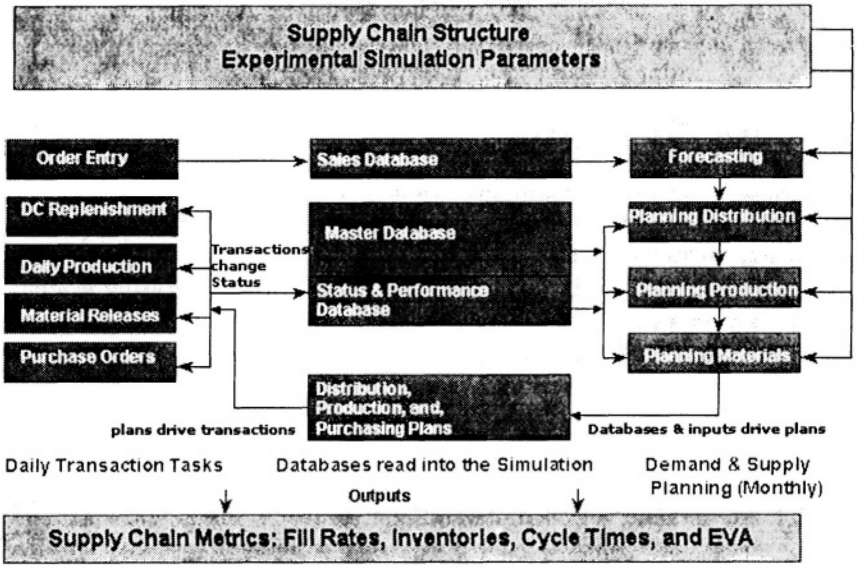


Figure 14.3. Simulation Overview

pages 269-276). An order is placed as soon as the inventory level reaches the reorder point. The supplying plant will then fulfill the order via its plant warehouse on the chosen mode of transport. If sufficient inventory is unavailable, the shipment is delayed until the appropriate lot-size (for the chosen mode) is available. The supplying plant warehouse, in turn, forecasts the orders it receives from the DCs. To simulate this, we use exactly the same procedure as in the DC forecasts, i.e., perturb the real orders with a pre-determined error component.

Under the CPFR type system, however, the plant (Denver in our case) has perfect visibility of the DC needs. This is achieved via a common planning database the firms share over the Internet. As in the ROP case, the DCs plan period-by-period product needs so that a pre-determined fill-rate is met. However, with CPFR, the plant and the DCs *jointly* plan the fulfillment of these needs. The plant-warehouse shipping schedule is achieved by first aggregating the DC needs in a given period and offsetting it by required transportation lead-time.

The production plan over the planning horizon is then computed as the quantity required to satisfy the warehouse shipping schedule and the safety stock requirements at the plant warehouse. Once the production plans are known, a standard MRP procedure is used determine raw material needs and shipping schedules from the suppliers. At the end of the "planning cycle" (beginning of every month) of the simulation, all plans–DC needs, plant warehouse shipping schedules, production plans, and raw material needs and shipping schedules are determined.

Once the plans are laid out for the coming year, the simulation starts simulating material flow every day according to these plans, collecting supply chain performance data every day of the simulation. All plans, from distribution, production, to raw material procurement are done using "averages," for example, when doing MRP, the mean lead-time is used to generate schedules and so on. Therefore, as the simulation proceeds with specific instances of forecasts, production and transport lead-times, the plans tend to deviate from reality. So the simulation updates the plans (for the horizon) at the end of month, i.e., the DC forecasts are updated, ROP or CPFR is performed, production and raw material plans are generated and so on. In our simulation model, after an initial warm-up period of three months, statistics are collected for three years to average random effects. Any product demand not satisfied is backordered except at the DCs, where it is accounted for as lost sales (and consequently fill-rates are collected). The actual service level at each of the DCs; the average inventory levels; transit statistics; and the financial performance at each entity in each of the four echelons are the key outputs of the simulation. In addition, we also compute the following overall supply chain measures:

Overall Customer Service level: Weighted average (by volume of product sold) of service levels at each DC. If ρ_i is the service level at DC_i and V_i is the volume of product flow at DC_i, then the overall service level is $\Sigma\rho_i V_i$

Economic Value Added (EVA): Profit - Cost of capital

Profit = Sales – Operating expenses (including inbound, production, outbound distribution)

Cost of capital = Cost of working capital (including inventory at various levels) + Cost of investment

Time through the supply chain: Weighted average by volume of either raw material, WIP, or finished goods of:

Time spent in transit from supplier to the plant 4 + Holding time in the plant (as raw material) + manufacturing time + holding time in the out-bound warehouse + transit time to the DCs + time spent in the

DCs + transit time to the customer.

For example, there are six plant warehouse-DC links (one for every DC). The "transit time to the DCs" is the weighted average of transit times from the plant warehouse to the DCs, weighted by the volume that was shipped on each of these links. This measures the time dimension in the supply chain or the "responsiveness" in the supply chain.

5. The Experiment

Our intent is to test the impact of collaborative initiatives on supply chain performance under a number of different operating conditions. To do so, we constructed a full factorial design to evaluate our hypotheses via the ANOVA procedure. We have chosen to change the following parameters as we believe that these are the most typical operating ranges of this supply chain:

a. Planning Options: CPFR or ROP.

b. Forecast Errors: "High" or "Low." High corresponds to an "a" parameter of 5; and Low to an "a" parameter of 3, with the parameter "b" estimated at 0.8. This represents the typical range of values observed for the forecast errors.

c. Service levels at the DCs: 90%, 95%, 99%. These are target fill-rates at the DCs. Effectively these are responsible for the appropriate levels of safety stock at the DC location.

d. Transport Modes: LTL, TL, TOFC/COFC, Rail Boxcar.

e. Levels of Safety Stock at the plant warehouse: 0.5, 1.0, 1.5 weeks of supply.

f. Average Levels of Demand: 45, 70, and 105 million pounds per year.

There are a total of 432 parameter combinations. Each combination was run at least 15 times, more if there were any outlying runs, for a total of 6522 runs.

We used SPSS v7.0 to analyze the outputs from these runs. To test our hypotheses, we use all the main effects and two significant interactions, Planning Options with Forecast Accuracy; and Planning Options with Fill Rates. We only report our key findings in this Chapter. For the interested reader, detailed analysis is available in Boone et al. (2000). ANOVA analysis shows that all the terms have a significant impact in determining fill rates, inventories, cycle times, and EVA.

Next, we proceeded to compare the mean levels of performance between the CPFR and the ROP control methods. Table 14.2 illustrates the relative performance of CPFR over the ROP method. For example, using the CPFR method yields on average fill-rates that are 1.8% higher relative to the ROP method. The standard error is small enough to make this percentage difference significant. This seems to provide evidence for the first hypothesis that the use of the CPFR method produces a higher fill-rate for the supply chain. For our data, the difference in fill-rates translates to additional sales anywhere from 675,000 pounds of the product (at the 45 million pounds level of demand) to 1,575,000 when the demand is 105 million pounds. Therefore the use of CPFR becomes very important for high volume items. When forecast errors are higher, CPFR yields fill-rates that are 3.3% higher relative to the ROP measure. Again, the standard errors are small making the difference significant. This suggests that the impact of CPFR on fill-rate increases as forecast errors increase, further confirming the fact that the biggest benefits of CPFR are when forecast errors are high.

Performance Measure	Performance Improvement from using CPFR	Standard Error
Supply Chain Time	-1.74%	< 0.5%
Supply Chain Inventory	-1.94%	< 0.3%
Observed Fill Rates	1.80%	< 0.1%
EVA	5.99%	< 1%

Table 14.2. Relative Performance of the CPFR Procedure

The CPFR procedure uses on average 1.94% less inventory (significant at 99% confidence) relative to the ROP, confirming hypothesis H2. Herein lies the biggest impact of information-sharing initiatives—they increase the fill-rates while reducing inventories. This is possible because the Denver plant has complete visibility of all the DC inventories. Therefore, Denver can plan shipping and production schedules more efficiently. Additionally, when forecast errors are higher, the decrease is 5.80%, indicating again the magnified impact of collaborative planning in high uncertainty.

CPFR procedure yields a cycle time that is on an average 1.74% less than that of the ROP procedure (significant at 99% confidence level), confirming hypothesis H3. The CPFR procedure warrants lesser inventory, and consequently a higher turnover ratio thus increasing the velocity of product flow across the supply chain.

Finally, the biggest impact of CPFR can be seen in the increase in shareholder wealth as measured by EVA. The CPFR procedure yields on average an EVA that is 5.99% higher than the ROP procedure (significant at 99% confidence), supporting H4. This difference can be attributed to the reductions in working capital and increase in revenues due to the CPFR procedure. At every entity, at each echelon in the supply chain, there is (i) a reduction of inventory, (ii) faster turnover rates leading to lower operating costs (recall that at the DCs and plants, the fixed and variable costs are a function of the volume), and finally (iii) higher revenues brought about by higher fill-rates. However the effect of CPFR on EVA may be exaggerated by our assumption that CPFR costs are 0.5% of the sales.

6. Summary and Conclusions

Our intent in this Chapter was to analyze in a systemic manner the benefits of information sharing mechanisms, specifically CPFR on four dimensions: fill-rates, supply chain inventory, supply chain cycle time, and shareholder value. We hypothesized that using CPFR increases margins and decreases working capital, consequently increasing fill-rates and EVA; and decreasing inventories and supply chain cycle time. We built an elaborate simulation of the supply chain, whose operations are adapted from a real company and uses real data. We then tested the impact of CPFR and the traditional ROP inventory planning method under a number of supply chain configurations and the subsequent analysis led to the following findings.

a. *CPFR increases fill-rates:* An increase in fill rates translates to a larger volume of product sold to the retail outlets thereby increasing the revenues and profit margins, due to lower costs. Additionally, the impact of CPFR is higher when the forecast errors are higher.

b. *CPFR decreases supply chain inventories:* At the plant-DC level, joint planning reduces any inventories that are used to buffer the added uncertainties that ROP systems warrant. This implies the plant will not have to inflate its production schedules to meet this excess inventory. This in turn impacts procurement of raw materials–plants with realistic schedules demand lower quantities and consequently hold lesser amounts of cycle inventories of raw materials in their warehouses. All this reduces the overall inventory level in the supply chain. Furthermore, the reduction in inventory is greater when the forecast errors are high. In certain industries with high uncertainty, like fashion goods, collaborative planning mechanisms can make a significant impact on reducing inventory levels.

c. *CPFR reduces supply chain cycle time:* The reduction of inventory in the entire pipeline increases the number of turns and hence speeds up the flow of the product from the raw material to the retail outlets. Hence CPFR leads to a time-compressed and a responsive supply chain.

d. *CPFR increases shareholder wealth:* High fill-rates and low inventories lead to higher margins and lower working capital, increasing EVA.

Although our research shows that collaborative planning has a substantial impact on the firm, we have not considered any implementation issues. We simulate the supply chain under the assumption that CPFR process can be easily implemented. This is not always the case as the CPFR system is not always quickly or easily adopted. The premise is that real-time data shared and planned together will benefit both parties. Several firms may not be willing to share sensitive sales or financial data. Furthermore, implementation of collaborative practices requires collaboration-support technology such as EDI, e-commerce applications, front-end and back-end application servers, and the appropriate databases to feed these collaborative-support technologies. It is easy to confuse the technology with 'collaborative planning', but the success of any partnership depends on the ability to use information, not having access to it. This would mean setting up joint teams between trading partners to analyze the data and to make joint decisions on demand, replenishment, and production in such a way that it benefits all involved. The role of the forecasting/logistics divisions of a firm will find itself in a different role–as partner and collaborator–but will make similar decisions: forecasting, safety stocks, production and shipping schedules, etc.

There is no simple formula to effectively implement CPFR initiatives in a firm. As Austin (1998) suggests, firms should use a three-pronged approach. First, a firm should evaluate the risk and rewards of a collaboration initiative. Much like the simulation described in this chapter, a firm can access the cost of implementation and the potential benefits of a collaborative initiative. Second, there is a need to reshape relationships between trading partners. Relationships between companies should move from just electronic transaction–be it over EDI or the Internet–to a more interactive one with the customer perspective in mind. Issues of trust and goodwill need to be addressed explicitly before the collaborative arrangements are undertaken. Third, as the nature of collaborative agreements change with time and with the improvement of technology, firms should make it a priority to reevaluate and execute newer and more effective collaborative agreements.

Future research can focus on the feasibility, costs, and benefits of CPFR and/or other information sharing agreements in other industries, especially in high technology and fashion industries, where compressed product life cycles and high uncertainties often lead to operating inefficiencies in the supply chain.

Acknowledgments

We would like to thank the participants of the Frank Batten Conference held at the College of William & Mary. This chapter is an abridged version of Boone, Ganeshan, and Stenger (2000), College of William & Mary Working Paper.

References

Austin, T., 1998. "The Personal Computer Supply Chain: Unlocking Hidden Value." In *Strategic Supply Chain Alignment,* Gower, Hampshire GU11 3HR, England, 188-208.

Boone, T., Ganeshan, R., and Stenger, A. J. 2000. "The Impact of Information Sharing Initiatives on Supply Chain Performance: A Simulation Study," College of William & Mary working paper.

Chen, F., Drezner, Z., Ryan, J. K., and Simchi-Levi, D., 1998. "The Bullwhip Effect: Managerial Insights on the Impact of Forecasting and Information on Variability on a Supply Chain." In *Quantitative Models for Supply Chain Management,* Tayur, Ganeshan, and Magazine (eds.), Kluwer Academic Press, Norwell, MA, 839-879.

Clark, T. and Hammond, J., 1997. "Reengineering Channel Reordering Processes to Improve Total Supply Chain Performance." *Production and Operations Management,* 6(3), 248-265.

Deuermeyer, B. and Schwarz, L. 1981. "A Model for the Analysis of System Service Level in Warehouse/Retailer Distribution Systems: The Identical Retailer Case." In *Studies in the Management Sciences: The Multi-Level Production/Inventory Control Systems,* L. Schwarz (ed.), North-Holland, Amsterdam, 163-193.

Forrester, J. 1961. *Industrial Dynamics,* MIT Press, Cambridge, MA.

Ganeshan, R., Jack, E., Magazine, M. J., and Stephens, P., 1998. "A Taxonomic Review of Supply Chain Management Research." In *Quantitative Models for Supply Chain Management,* Tayur, Ganeshan, and Magazine (eds.), Kluwer Academic Press, Norwell, MA, 839-879.

Lee, H., Padmanabhan, V., and Whang, S. (1997a). "The Bullwhip Effect in Supply Chains." *Sloan Management Review,* 38, 93-102.

Lee, H., Padmanabhan, and Whang, S. (1997b). "Information Distortion in a Supply Chain: the Bullwhip Effect." *Management Science,* 43, 546-558.

Long, M., 1999. Personal Communication

Sharp, D. and Hill, R., 1998. "ECR: From Harmful Competition to Winning Collaboration." In *Strategic Supply Chain Alignment,* Gower, Hampshire GU11 3HR, England, 104-122.

Sridharan, S. and Berry, W. 1990. "Freezing the Master Production Schedule Under Demand Uncertainty." *Decision Sciences,* 21(1), 97-120.

Stenger, A. J. 1994. "Inventory Decision Framework." In *The Logistics Handbook,* Robeson, J. F. and W. C. Copucino (eds.), The Free Press, New York, NY, 352-371.

Tayur, S., Ganeshan, R., and M. J. Magazine, 1998 *Quantitative Models for Supply Chain Management.* Kluwer Academic Press, Norwell, MA.

White, A., 1999. "The Value Equation: Value Chain Management, Collaboration, and the Internet." White Paper, Logility Inc., 470 E. Paces Ferry Rd., Atlanta, GA 30305.

Applied Optimization

36. G. Di Pillo and F. Giannessi (eds.): *Nonlinear Optimization and Related Topics.* 2000 ISBN 0-7923-6109-1
37. V. Tsurkov: *Hierarchical Optimization and Mathematical Physics.* 2000 ISBN 0-7923-6175-X
38. C. Zopounidis and M. Doumpos: *Intelligent Decision Aiding Systems Based on Multiple Criteria for Financial Engineering.* 2000 ISBN 0-7923-6273-X
39. X. Yang, A.I. Mees, M. Fisher and L.Jennings (eds.): *Progress in Optimization.* Contributions from Australasia. 2000 ISBN 0-7923-6286-1
40. D. Butnariu and A.N. lusem: *Totally Convex Functions for Fixed Points Computation and Infinite Dimensional Optimization.* 2000 ISBN 0-7923-6287-X
41. J. Mockus: *A Set of Examples of Global and Discrete Optimization.* Applications of Bayesian Heuristic Approach. 2000 ISBN 0-7923-6359-0
42. H. Neunzert and A.H. Siddiqi: *Topics in Industrial Mathematics.* Case Studies and Related Mathematical Methods. 2000 ISBN 0-7923-6417-1
43. K. Kogan and E. Khmelnitsky: *Scheduling: Control-Based Theory and Polynomial-Time Algorithms.* 2000 ISBN 0-7923-6486-4
44. E. Triantaphyllou: *Multi-Criteria Decision Making Methods.* A Comparative Study. 2000 ISBN 0-7923-6607-7
45. S.H. Zanakis, G. Doukidis and C. Zopounidis (eds.): *Decision Making: Recent Developments and Worldwide Applications.* 2000 ISBN 0-7923-6621 -2
46. G.E. Stavroulakis: *Inverse and Crack Identification Problems in Engineering Mechanics.* 2000 ISBN 0-7923-6690-5
47. A. Rubinov and B. Glover (eds.): *Optimization and Related Topics.* 2001 ISBN 0-7923-6732-4
48. M. Pursula and J. Niittymäki (eds.): *Mathematical Methods on Optimization in Transportation Systems.* 2000 ISBN 0-7923-6774-X
49. E. Cascetta: *Transportation Systems Engineering: Theory and Methods.* 2001 ISBN 0-7923-6792-8
50. M.C. Ferris, O.L. Mangasarian and J.-S. Pang (eds.): *Complementarity: Applications, Algorithms and Extensions.* 2001 ISBN 0-7923-6816-9
51. V. Tsurkov: *Large-scale Optimization – Problems and Methods.* 2001 ISBN 0-7923-6817-7
52. X. Yang, K.L. Teo and L. Caccetta (eds.): *Optimization Methods and Applications.* 2001 ISBN 0-7923-6866-5
53. S.M. Stefanov: *Separable Programming Theory and Methods.* 2001 ISBN 0-7923-6882-7

Applied Optimization

54. S.P. Uryasev and P.M. Pardalos (eds.): *Stochastic Optimization: Algorithms and Applications.* 2001 ISBN 0-7923-6951-3
55. J. Gil-Aluja (ed.): *Handbook of Management under Uncertainty.* 2001 ISBN 0-7923-7025-2
56. B.-N. Vo, A. Cantoni and K.L. Teo: *Filter Design with Time Domain Mask Constraints: Theory and Applications.* 2001 ISBN 0-7923-7138-0
57. S. Zlobec: *Stable Parametric Programming.* 2001 ISBN 0-7923-7139-9
58. M.G. Nicholls, S. Clarke and B. Lehaney (eds.): *Mixed-Mode Modelling: Mixing Methodologies for Organisational Intervention.* 2001 ISBN 0-7923-7151-8
59. F. Giannessi, P.M. Pardalos and T. Rapcsák (eds.): *Optimization Theory. Recent Developments from Mátraháza.* 2001 ISBN 1-4020-0009-X
60. K.M. Hangos, R. Lakner and M. Gerzson: *Intelligent Control Systems.* An Introduction with Examples. 2001 ISBN 1-4020-0134-7

KLUWER ACADEMIC PUBLISHERS – DORDRECHT / BOSTON / LONDON