International Series in
Operations Research & Management Science

Tsan-Ming Choi *Editor*

Handbook of Newsvendor Problems

Models, Extensions and Applications

Springer

International Series in Operations Research & Management Science

Volume 176

For further volumes:
http://www.springer.com/series/6161

Tsan-Ming Choi
Editor

Handbook of Newsvendor Problems

Models, Extensions and Applications

Springer

Editor
Tsan-Ming Choi
Business Division, Institute of Textiles
and Clothing
The Hong Kong Polytechnic University
Hung Hom, Kowloon
Hong Kong SAR

ISSN 0884-8289
ISBN 978-1-4614-3599-0 ISBN 978-1-4614-3600-3 (eBook)
DOI 10.1007/978-1-4614-3600-3
Springer New York Heidelberg Dordrecht London

Library of Congress Control Number: 2012939061

Printed on acid-free paper

Springer is part of Springer Science+Business Media (www.springer.com)

Preface

Inventory management is a critical factor which accounts for the success or failure of modern businesses in nearly all kinds of industries. As a fundamental problem in stochastic inventory control, the newsvendor problem has been studied since the eighteenth century in the economic literature and it has been widely used to analyze supply chains with fashionable and perishable products. Starting from the 1950s, newsvendor problem has been extensively studied in operations research and extended to model a large variety of real-life problems. The simplest and most elementary version of the newsvendor problem is an optimal inventory stocking problem in which a newsvendor needs to decide how much newspaper to order for the future demand, where the future demand is uncertain and follows a stationary distribution. The newspaper becomes obsolescent at the end of the day and the newspaper's cost–revenue structure is known. The newsvendor then determines its optimal order quantity by either maximizing the expected profit or minimizing the expected cost and an analytical closed-form solution exists. This classical newsvendor problem has been extended in many different ways and nowadays, a search in major research portals will find at least thousands of papers related to this problem and every year, tens to hundreds of related papers are still being published in major journals on operations research and management science. Despite the abundance of both classical and new research results, there is an absence of a comprehensive reference source that provides the state-of-the-art findings on both theoretical and applied research on the newsvendor problem. As a result, I organize this Springer's handbook with a goal of consolidating many latest research findings and applications of the newsvendor problem into an edited volume. I believe that this handbook will be a pioneering text focusing on the newsvendor problem.

The handbook is organized into two parts, namely (1) models and extensions and (2) applications of the newsvendor problem. I am very pleased to see that this handbook has generated a lot of new research results with valuable insights into the following topics:

- A Timely Review on the Multi-Product Newsvendor Problem
- A Multi-Product Risk-Averse Newsvendor with Law Invariant Coherent Measures of Risk

- A Copula Approach to Inventory Pooling Problems with Newsvendor Products
- Repeated Newsvendor Games with Transshipments
- Cooperative Newsvendor Games
- An Economic Interpretation for the Price-Setting Newsvendor Problem
- Newsvendor Models with Alternative Risk Preferences within Expected
- Utility Theory and Prospect Theory Frameworks
- Newsvendor Problems with VaR and CVaR Consideration
- A Two-Period Newsvendor Problem for Closed-Loop Supply Chain Analysis
- The Remanufacturing Newsvendor Problem
- Inventory Centralization in a Newsvendor Setting when Shortage Costs Differ
- Production Planning on an Unreliable Machine for Multiple Items Subject to Stochastic Demand
- Analysis of the Newsvendor Problem Under Carbon Emissions Policies
- Optimal Decisions of the Manufacturer and Distributor in a Fresh Product Supply Chain Involving LongDistance Transportation
- A Newsvendor Perspective on Profit Target Setting for Multiple Divisions
- A Portfolio Approach to Multi-Product Newsvendor Problem with Budget Constraint

I would like to take this opportunity to show my hearty gratitude to Fred Hillier and Matthew Amboy for their kind support and advice along the course of carrying out this book project. I sincerely thank all the authors who have contributed their decent research to this handbook. I am indebted to the anonymous reviewers who reviewed the manuscripts and provided me with constructive review comments. I also acknowledge the editorial assistance of my research students Dr Pui-Sze Chow and Ms Hau-Ling Chan and the funding support of the Research Grants Council of Hong Kong under grant number PolyU 5424/11H (General Research Fund). Last but not least, I am grateful to my family, colleagues, and students, who have been supporting me during the development of this important research handbook.

Kowloon, Hong Kong Tsan-Ming Choi

Contents

Part II Applications

Contributors

Miguel Ampudia Department of Economics, Boston University, Boston, MA, USA

Burcu Aydın HP Labs, Palo Alto, CA, USA

Niv Ben-Zvi Department of Industrial Engineering, Tel-Aviv University, Tel-Aviv, Israel

Xiaoqiang Cai Department of Systems Engineering and Engineering Management, The Chinese University of Hong Kong, Shatin, N.T., Hong Kong

Jian Chen Research Center for Contemporary Management, Key Research Institute of Humanities and Social Sciences at Universities, School of Economics and Management, Tsinghua University, Beijing, China

Sungyong Choi Department of Business Administration, College of Government and Business, Yonsei University, Wonju, Republic of Korea

Matthew J. Drake Palumbo-Donahue Schools of Business, Duquesne University, Pittsburgh, PA, USA

Yigal Gerchak Department of Industrial Engineering, Tel-Aviv University, Tel-Aviv, Israel

Kemal Guler HP Labs, Palo Alto, CA, USA

Lan Guo Accounting, School of Business and Economics, Wilfrid Laurier University, Waterloo, Ontario, Canada

Zhongsheng Hua School of Management, University of Science and Technology of China, Hefei, Anhui, China

Xiao Huang John Molson School of Business, Concordia University, Montr'eal, Qu'ebec, Canada

Werner Jammernegg Department of Information Systems and Operations, WU Vienna University of Economics and Business, Vienna, Austria

Enis Kayış HP Labs, Palo Alto, CA, USA

Peter Kischka Department of Business Statistics, Friedrich Schiller University Jena, Jena, Germany

David Kletter Booz Allen Hamilton, New York, NY, USA

Gernot Lechner Institute of Production and Operations Management, University of Graz, Universitätsstrasse, Graz, Austria

Mingming Leng Department of Computing and Decision Sciences, Faculty of Business, Lingnan University, Tuen Mun, Hong Kong

Luigi Montrucchio Dipartimento di Statistica e Matematica Applicata and Collegio Carlo Alberto, Universita di Torino, Corso Unione Sovietica, Torino, Italy

Henk Norde Center and Department of Econometrics and OR, Tilburg University, Tilburg, Netherlands

Ulaş Özen Alcatel-Lucent Bell Labs Ireland, Dublin, Ireland

David W. Pentico Palumbo-Donahue Schools of Business, Duquesne University, Pittsburgh, PA, USA

Marc Reimann Institute of Production and Operations Management, University of Graz, Universitätsstrasse, Graz, Austria

Michael A. Salinger School of Management, Boston University, Boston, MA, USA

Marco Scarsini Dipartimento di Scienze Economiche e Aziendali, LUISS, Roma, Italy

Chunming (Victor) Shi Operations and Decision Sciences, School of Business and Economics, Wilfrid Laurier University, Waterloo, Ontario, Canada

Marco Slikker School of Industrial Engineering, Technische Universiteit Eindhoven, Paviljoen E13, Eindhoven, Netherlands

Jingpu Song International Business School, Shanghai Institute of Foreign Trade, Songjiang, Shanghai, China

Greys Sošić Marshall School of Business, University of Southern California, Los Angeles, CA, USA

Robert P. Sroufe Palumbo-Donahue Schools of Business, Duquesne University, Pittsburgh, PA, USA

Yinliang Tan Department of Information Systems and Operations Management, University of Florida, Gainesville, FL, USA

Nazli Turken Department of Information Systems and Operations Management, University of Florida, Gainesville, FL, USA

Asoo J. Vakharia Department of Information Systems and Operations Management, University of Florida, Gainesville, FL, USA

Charles X. Wang School of Management, State University of New York at Buffalo, Buffalo, NY, USA

Lan Wang Department of Information Systems and Operations Management, University of Florida, Gainesville, FL, USA

Ruoxuan Wang Department of Information Systems and Operations Management, University of Florida, Gainesville, FL, USA

Scott Webster Whitman School of Management, Syracuse University, Syracuse, NY, USA

Yongbo Xiao Research Center for Contemporary Management, Key Research Institute of Humanities and Social Sciences at Universities, School of Economics and Management, Tsinghua University, Beijing, China

Xiaolin Xu School of Business, Nanjing University, Nanjing, China

Arda Yenipazarli Department of Information Systems and Operations Management, University of Florida, Gainesville, FL, USA

Gang Yu Department of Management Science and Information Systems, The University of Texas at Austin, Austin, TX, USA

Bin Zhang Lingnan College, Sun Yat-Sen University, Guangzhou, China

Sidong Zhang Chinesisch-Deutsches Hochschulkolleg, Tongji University, Shanghai, China

Part I
Models and Extensions

Chapter 1
The Multi-product Newsvendor Problem: Review, Extensions, and Directions for Future Research

Nazli Turken, Yinliang Tan, Asoo J. Vakharia, Lan Wang, Ruoxuan Wang, and Arda Yenipazarli

Abstract In this paper, we review the contributions to date for the multi-product newsvendor problem (MPNP). Our focus is on the current literature concerning the mathematical models and the solution methods for the multi-item newsvendor problems with single or multiple constraints, as well as the effects of substitute and complementary products on the stocking decisions and expected profits. We present some extensions to the current work for a stylized setting assuming two products and conclude with directions for future research.

Keywords Multi-product newsvendor • Complementary products • Effects of substitute • Single constraint • Multiple constraints • Future research

1.1 Introduction

The single-item newsvendor problem is one of the classical problems in the literature on inventory management (Arrow et al. 1951; Silver et al. 1998) and the reader interested in a comprehensive review of extant contributions for analyzing the problem is referred to Qin et al. (2011). In this paper, we focus on the multi-product newsvendor problem (MPNP) which can be framed as follows. At the beginning of a single period, a buyer is interested in determining a stocking policy (Q_i) for product i ($i = 1, \ldots, n$) to satisfy total customer demand for each product. For each product i, the customer demand is assumed to be stochastic and characterized by a random variable x_i with the probability density function $f_i(\cdot)$ and cumulative distribution

N. Turken • Y. Tan • A.J. Vakharia (✉) • L. Wang • R. Wang • A. Yenipazarli
Department of Information Systems & Operations Management, University of Florida, Gainesville, FL 32611–7169, USA
e-mail: nazli.z.turken@warrington.ufl.edu; tanyinliang@ufl.edu; asoov@ufl.edu; lan.wang@warrington.ufl.edu; rxwang@ufl.edu; arda.yenipazarli@warrington.ufl.edu

T.-M. Choi (ed.), *Handbook of Newsvendor Problems*, International Series in Operations Research & Management Science 176, DOI 10.1007/978-1-4614-3600-3_1,

function $F_i(\cdot)$. The quantity Q_i is purchased by the buyer for a fixed price per unit v_i. Assuming no capacity restrictions on the purchase quantity and zero purchasing lead time, an order placed by the buyer with the supplier at the beginning of a period is immediately filled. Sales of the product occur during (or at the end of) the single period and for each product i: (a) if $Q_i \geq x_i$, then $Q_i - x_i$ units which are left over at the end of the period are salvaged for a per unit revenue of g_i[1]; and (b) if $Q_i < x_i$, then $x_i - Q_i$ units which represent "lost" sales cost B_i per unit. Assuming a fixed market price of p_i, then the actual end of period profit for the buyer stemming from the sales of each product i is:

$$\pi_i(Q_i, x_i) = \begin{cases} p_i x_i - v_i Q_i + g_i(Q_i - x_i) & \text{if } Q_i \geq x_i \\ p_i Q_i - v_i Q_i - B_i(x_i - Q_i) & \text{if } Q_i < x_i \end{cases}. \tag{1.1}$$

Since the buyer cannot observe the actual end-of-period profit when making his decision at the beginning of the period, the traditional approach to analyze the problem is based on assuming a risk neutral buyer who makes the optimal quantity decision at the beginning of the period by maximizing total expected profits. These profits are:

$$\begin{aligned} E[\pi_i(Q_i)] &= \int_0^{Q_i} [p_i x_i - v_i Q_i + g_i(Q_i - x_i)] f_i(x_i) \mathrm{d}x_i \\ &\quad + \int_{Q_i}^{\infty} [p_i Q_i - v_i Q_i - B_i(x_i - Q_i)] f_i(x_i) \mathrm{d}x_i \\ &= (p_i - g_i)\mu_i - (v_i - g_i)Q_i - (p_i - g_i + B_i) ES_i(Q_i), \end{aligned} \tag{1.2}$$

where $E[\cdot]$ is the expectation operator, μ_i is the mean demand for product i, and $ES_i(Q_i)$ represents the expected units short assuming Q_i units are stocked and can be determined as $\int_{Q_i}^{\infty} (x_i - Q_i) f_i(x_i) \mathrm{d}x_i$. Based on this, the buyer's expected profits for the MPNP are:

$$\begin{aligned} E[\Pi(Q_1,..,Q_n)] &= \sum_{i=1}^{n} E[\pi_i(Q_i)] \\ &= \sum_{i=1}^{n} [(p_i - g_i)\mu_i - (v_i - g_i)Q_i - (p_i - g_i + B_i) ES_i(Q_i)]. \end{aligned} \tag{1.3}$$

Note that (1.1) is separable in each product i. Given that (1.2) is strictly concave in Q, it follows that the first order conditions (FOCs) for optimizing (1.2) are necessary and sufficient to determine the optimal value of Q_i. Based on this, the optimal stocking quantity for each product i (Q_i^*) is set such that:

$$F_i(Q_i^*) = \frac{p_i - v_i + B_i}{p_i - g_i + B_i} \tag{1.4}$$

[1]For obvious reasons, it is generally assumed that $g_i < v_i$.

and the corresponding total profit for the buyer is:

$$E[\Pi(Q_1^*, Q_2^*, .., Q_n^*)] = \sum_{i=1}^{n} \left[(p_i - g_i)\mu_i - (p_i - g_i + B_i) \int_{Q_i^*}^{\infty} x_i f_i(x) \mathrm{d}x_i \right]. \tag{1.5}$$

Our focus in this chapter is on reviewing and extending the current literature related to the MPNP. To start with, Sect. 1.2 reviews the current contributions for analyzing the MPNP with one or more stocking constraints. In Sect. 1.3, we review prior work which focuses on product substitutability in the context of the MPNP. Extensions of the MPNP for complementary/substitute products are described in Sect. 1.4 and finally, in Sect. 1.5, we conclude with directions for future research.

1.2 Buyer Stocking Constraints

1.2.1 Single Constraint

The general problem in this setting is to optimize the total profit in (1.5) subject to the following constraint:

$$\sum_{i=1}^{n} s_i Q_i \leq S, \tag{1.6}$$

where s_i is the storage space or the resource coefficient required per unit of product i and S is the total available storage space or resource. Since (1.3) is strictly and jointly concave in the decision variables, Q_is, and the constraints are linear, one approach to solving this problem would be to start with the solution to the unconstrained MPNP and substitute this solution in the constraint. If the constraint is not violated, then we have an optimal solution. Of course, the issue that needs to be considered is how to solve the problem when the constraint is violated with the solution to the unconstrained MPNP.

Hadley and Whitin (1963) proposed a Lagrange multiplier technique and a dynamic programming solution procedure for finding the optimal order quantity in this setting. The Lagrangian for this context is:

$$L(Q_i, \lambda) = E[\pi(Q_1, ., Q_n)] - \lambda \left(\sum_{i=1}^{n} s_i Q_i - S \right). \tag{1.7}$$

Since (1.7) is strictly and jointly concave in the decision variables, the FOCs are necessary and sufficient to obtain an optimal solution. These FOCs are:

$$\frac{\partial L}{\partial \lambda} = [g_i - v_i] F(Q_i^*) + [1 - F(Q_i^*)][p_i - v_i + B_i] - \lambda s_i, \tag{1.8}$$

$$\frac{\partial L}{\partial Q_i} = \sum_{i=1}^{n} s_i Q_i - S. \tag{1.9}$$

Setting (1.8) equal to 0, the optimal stocking quantity for each product Q_i^* is:

$$Q_i^* = F_i^{-1}\Big(\frac{p_i - v_i - \lambda s_i + B_i}{p_i - g_i + B_i}\Big), \tag{1.10}$$

where $\lambda \geq 0$.

In some practical situations, the optimal Q_i^* will tend to be very small and any attempt to use the above procedure and round the results (to obtain integer values of Q_i^*) could lead to considerable deviations from optimality. To handle this situation, the authors propose a dynamic programming-based procedure but of course, this method is not easily applicable when the number of products (n) is significantly large.

Nahmias and Schmidt (1984) introduced several heuristic methods to solve the MPNP with a single constraint where the lagrange multiplier, λ, is not easy to evaluate. They also included an interest rate, I,which is used in determining the carrying charge per period for the average inventory. Hence, the expected profit including the interest rate can be shown as:

$$E[\pi_i(Q_i)] = (p_i + 0.5Iv_i - g_i)\mu_i - [(1-I)v_i - g_i]Q_i - (p_i + 0.5Ic_i - g_i + B_i)ES_i(Q_i). \tag{1.11}$$

The optimal quantity then becomes:

$$Q_i^* = F^{-1}\left[\frac{p_i - (1+0.5I)v_i + B_i - \lambda s_i}{p_i + 0.5Iv_i - g_i + B_i}\right]. \tag{1.12}$$

Guessing an appropriate value of λ, computing the corresponding values of Q_i and subsequently adjusting the value of λ depending on (1.12) is very time consuming. Thus, four different heuristic methods were introduced. Heuristic 1 finds the solution to the unconstrained problem and adjusts these values until the constraint is satisfied. In heuristic 2, the critical point of the demand distribution is scaled to fit the given volume. Finally, heuristics 3 and 4 are proposed based on the Taylor expansion series of $t_i(\lambda) = \Phi^{-1}(a_i - b_i\lambda)$ and the corresponding λs could be calculated as follows:

$$\lambda_i = \frac{\sum_{i=1}^{n} \mu_i s_i + \sqrt{2\pi}\sum_{i=1}^{n} s_i(a_i - 0.5)\sigma_i - S}{\sqrt{2\pi}\sum_{i=1}^{n} b_i \sigma_i v_i}. \tag{1.13}$$

The procedures listed in this paper are mostly useful for the continuous values of Q_is and are thus appropriate for moderate-to-high demand items.

Lau and Lau (1996) were among the first to observe that using Hadley and Whitin's approach may lead to infeasible(negative) order quantities for some of the considered products when the constraint is tight. They based their work on the classical expected cost minimization problem that was introduced by Hadley and Whitin, where $(v_i - g_i)$ is the unit overage cost and $(p_i - v_i + B_i)$ is the unit underage cost. By rearranging terms of (1.8), we can find the expected net benefit of the marginal unit of product i ($EBMU_i$) at Q_i as:

$$EBMU_i = [p_i - v_i + B_i][1 - F(Q_i^*)] - [v_i - g_i]F(Q_i^*). \tag{1.14}$$

Note that $EBMU/s_i$ is analogous to the λ_i. Lau and Lau introduced a procedure to handle distributions with strictly positive lower bounds as well as distributions with long left tails.

Abdel-Malek et al. (2004) developed the exact solution formulae for uniformly distributed demand and presented a generic iterative method (GIM) when the demand distribution is general. The author considered the total budget as the resource constraint ($\sum_{i=1}^{n} v_i Q_i \leq B_G$). Different from most of the work in the literature, the author assumes there is a leftover cost (disposal fee), where a salvage value is usually considered. In general, if the budget is abundant, the problem could be solved by the unconstrained solution, yet if the budget is tight, we need to apply the Lagrangian-based approach to solve the problem. The value of λ is crucial to solve the problem and the author discusses how to address this under specific and general demand distributions. The formula for λ when the demand is uniformly distributed between a_i and b_i can be written as:

$$\lambda_u = \frac{\sum_{i=1}^{N}(c_i x_i^*) - B_G}{\sum_{i=1}^{N}(b_i - a_i)(v_i^2)/(p_i + h_i)}, \tag{1.15}$$

where h_i is the holding cost. The closed-form expression when the demand is exponential:

$$\lambda_e^{(1)} = \frac{\sum_{i=1}^{N}(c_i x_i^*) - B_G}{\sum_{i=1}^{N}(\mu_i - v_i^2)/(v_i + h_i)}. \tag{1.16}$$

The proposed GIM finds the optimum under uniform distribution and near optimum for other general distributions. GIM first finds the solution without the constraint and checks whether the constraint is satisfied. If the constraint is satisfied, the solution is optimal, if not, a solution that satisfies the constraint is found. Next, the error is estimated, if this is at an acceptable level, the optimal solution is found. As an extension to this paper, Abdel-Malek and Montanari (2005a) also defined the thresholds to help the decision maker in recognizing the tightness of the budget constraint, which can avoid infeasible order quantities by removing products with low marginal utilities. The following equations determine the thresholds depending on the available budget and demand patterns:

Threshold 1:

$$B_G^{(1)} = \sum_{i=1}^{n} v_i F_i^{(-1)}\left(\frac{p_i - v_i + B_i}{p_i - g_i + B_i}\right) = \sum_{i=1}^{n} v_i Q_i^*. \tag{1.17}$$

Threshold 2:

$$B_G^{(2)} = \sum_{i=1}^{n} v_i F_i^{(-1)} \left(\frac{p_i - (\theta^- + 1)v_i + B_i}{p_i - g_i + B_i} \right), \tag{1.18}$$

where $\theta^- = \min_{(i=1,.n)}(\theta_i)$ and notice that θ_i is the marginal utility at the lower limit of the feasible amount of the product to be ordered and could be calculated as follows:

$$\theta_i = \frac{p_i + B_i - (p_i - g_i + B_i)F_i(0)}{v_i} - 1. \tag{1.19}$$

Once the thresholds are defined, the solution procedure for each of the resulting cases can be implemented as shown in the following.

Case 1 $B_G^{(1)} \leq B_G$. In this case, the budget is abundant and the budget constraint is redundant, so we can obtain the optimal solution from the unconstrained problem.

$$F_i(Q_i^*) = \frac{(p_i - v_i + B_i)}{(p_i - g_i + B_i)}. \tag{1.20}$$

Case 2 $B_G^{(2)} \leq B_G < B_G^{(1)}$. In this case, we can relax the nonnegativity constraint and use the Lagrange method to get the optimal solutions.

$$F_i(Q_i^*) = \frac{p_i - (\theta + 1)v_i + B_i}{p_i - g_i + B_i}. \tag{1.21}$$

Case 3 $B_G < B_G^{(2)}$. In this case, as mentioned before, the nonnegativity constraints should be added to the model to avoid the infeasible solution. Furthermore, one or more products will have an order quantity of zero.

To determine the optimal order quantities, one needs to compute the marginal utilities of each product by using (1.19) first and rank them in ascending order. Begin with excluding the product (set order quantity to zero) from the top of the list and continue the exclusion process until the updated budget threshold is less than the previous budget. $B_{(G,i)}$ is the updated budget threshold as well as the lower bound of budget required for including item i in the list, which is expressed by:

$$B_{G,i} = \sum_{i=1}^{n'} v_i F_i^{-1} \left(\frac{p_i - (\theta_i + 1)v_i + B_i}{p_i - g_i + B_i} \right) < B_G. \tag{1.22}$$

n' is the updated number of items on the list. Once this point is reached, the problem becomes tractable again and we can apply the Lagrangian method without nonnegativity constraint to get the optimal solutions.

Several researchers incorporated nonnegativity constraints on the decision variables in their approaches. Erlebacher (2000) developed optimal and heuristic solutions for the classical problem. The first optimal solution refers to the event where each item has a similar cost structure and the demand for each item is from a similar distribution. The second case is when the demand for each item follows a uniform distribution. The first heuristic (H1) is optimal when all of the items have a similar cost structure and similar shaped demand distribution and requires only the mean and variance of each demand distribution and the cost data. The second heuristic (H2) is optimal when the demand is uniformly distributed for each item. The third heuristic (H3) is a modification of (H1) to account for general cost structures based on the form of (H2). The authors use computational experiments to show that (H2) is the most effective one, especially at higher levels of capacity.

Zhang et al. (2009) developed a binary search method to obtain the optimal solution. They defined the marginal benefit function as $r_i(x_i) = (v_i - B_i + \frac{(g_i+B_i)F_i(x_i)}{v_i})$, where $r_i(x_i)$ is a nondecreasing function of x_i, when $x_i \geq 0$ and its inverse is a strictly increasing function of r_i when $1 - \frac{B_i}{v_i} \leq r_i(x_i) < 0$. The authors find that the optimal solution to the constrained problem is the same as the unconstrained optimal solution when the budget constraint is not binding and is less than the unconstrained optimal solution when the budget constraint is binding. If there are nonzero optimal solutions, their marginal benefits should equal each other. When the budget constraint is binding, the optimal solution is $x_i^{(**)}$, and $r^{(**)} = r_i(x_i)$ is the marginal benefit at $x_i^{(**)}$. Zhang and Hua showed that $1 - \frac{B_i}{v_i} \leq r^{(**)} < 0$ and $r^{(**)}$ can be found using a binary search between these values. The algorithm they developed first finds the solution to the unconstrained problem and assesses whether the optimal value leads to a binding budget constraint. If this solution does not satisfy the condition, a binary search procedure is applied. This algorithm can provide an optimal or a near optimal solution to MPNP under any general demand distribution and it can also provide a good approximate solution under discrete demand distributions.

Zhang and Du (2010) studied the MPNP with a capacity constraint, where the products can be outsourced to an external facility at a higher cost. They considered zero-lead time (ZO) and nonzero lead time (NO) strategies. In ZO strategy, the manufacturer makes the decision for the in-house production quantity in the first period, and in the second period, after the demand is realized, the manufacturer outsources the remaining demand with zero lead time. There are no lost sales in this case. In the NO strategy, the manufacturer makes the decision for the in-house production quantity and the outsourcing quantity in the first period. In this strategy, if the demand exceeds the in-house production and outsourcing, there will be backorders or lost sales. The NO strategy assumes that there is no difference in arrival times of the products whether they are outsourced or produced in-house or if a time difference exists, it is assumed that there is no cost to receiving the product earlier than required.

It is assumed that each product has a deterministic production capacity and a random demand. The demand distributions are approximated to exclude the negative values allowing the following assumptions: $F_i(x) = 0$ for all $x < 0$, and $F_i(0) \geq 0$. The expected profit function for the ZO strategy consists of the revenue of product i, salvage value of the excess of product i, less the outsourcing and in-house production cost of product i. This model can be viewed as a parameter-adjusted single-constraint newsvendor model, and can be solved using the methods developed by Zhang et al. (2009). Similarly, the expected profit function of the NO strategy can be written as $\pi_2 = \sum_{i=1}^{n}[(p_i - v_i)Y_{i+}(p_i - d_i)Z_i - (p_i - g_i)\int_0^{Y_i+Z_i} F_i(x_i)\mathrm{d}x_i$, where d_i is the cost of outsourcing one unit, Y_i and Z_i are the decision variables for in-house production and outsourcing, respectively. By analyzing the partial derivatives and the KKT conditions, it is evident that the constrained optimal solution for in-house production will always be smaller than the unconstrained optimal solution for in-house production when there is no outsourcing ($Y_i^* \leq \widetilde{Y}$). The optimal outsourced quantity will also be less than the unconstrained problem solution and the maximum value it can take is $\widetilde{Z} - Y^*$. If the unconstrained optimal in-house production quantity does not exceed the available capacity, everything will be produced in-house. If the unconstrained optimal in-house production quantity exceeds the available capacity, the limited capacity must be fully utilized. Finally, the optimal solutions can be designed in a way that there exists only one product that utilizes both sources of production,and for every other product only one source of production is used. The results are that ZO strategy outperforms the NO strategy when outsourcing costs are equal and managers should try to find a ZO option with low implementing costs to achieve the maximum profit.

Moon and Silver (2000) presented dynamic programming procedures for MPNP, where the budget is represented as the total value of the replenishment quantities. In this paper, the decision variable is the order-up-to level, S_i. There is an inventory level of I_i at the beginning of period i, a fixed ordering cost of A_i, and a variable cost of $G_i^F(S_i^*)$. Initially, it is assumed that there is enough budget to permit each item to be ordered at its optimal. The authors formulate the problem as a minimization of fixed and variable costs $C_i^F(S_i)$, and decide the ordering rule to be: order up to S_i^*, if $I_i < s_i^*$ where $G_i^F(s_i^*) = A_i + G_i^F(S_i^*)$. Moon and Silver, then introduced a restricted budget W and developed a dynamic program to find the optimal order-up-to level. This dynamic program first tries to solve the single period model with a fixed ordering cost for each item separately and defines **P** to be the set of items that are profitable to order and reaches to an optimum when the ordering cost is within the budget. This solution method will become time consuming if the number of items or the number of budget constraints are high. Hence, the authors developed two heuristic algorithms. The first one, is a greedy allocation algorithm. At each step, the algorithm reduces the budget until a feasible solution is reached and any remaining budget is filled in a reverse greedy manner. The second algorithm, a two-stage heuristic, tries to assign the budget proportionally to the items in **P**.

The authors also considered the distribution-free model and assumed that the distribution of the demand belongs to the class **F** cumulative distribution functions with mean μ_j and variance σ_j^2. This approach requires finding the most unfavorable distribution in **F** for each S. Then, the objective function becomes $\min_{(S_1,\ldots,S_m)} \max_{(F\in F)} \sum_{i=1}^{m} C_i^F(S_i)$. The authors rewrite the cost function as $\sum_{i=1}^{m} G_i^W(S_i) + A_{(i)} 1_{(S_i > I_i)}$ using the proposition from Gallego and Moon (1993) indicating that a distribution satisfying $E[D_i - S_i]^+ \leq \frac{1}{2}\{\sqrt{[(\sigma_i^2) + (S_i - \mu_i^2)^2]} - (S_i - \mu_i)\}$ can always be found. In this cost function,

W denotes a worst case distribution function of the demand. The optimal solution can be found through backward recursive equations. The use of this distribution-free solution is justified when the expected value of additional information $(EVAI) = \sum_{i=1}^{m} C_i^N(S_i^W) - \sum_{i=1}^{m} C_i^N(S_i^N)$ is low. They mentioned two heuristics can also be modified to solve the distribution-free approach; however, this has not been studied in this paper.

Shao and Ji (2006) studied the multi-item newsvendor problem where the demand is fuzzy. They defined the profit for product i to be:

$$f(x_i, x_i(\xi_i)) = \begin{cases} (p_i - v_i)Q_i & \text{if } Q_i \le x_i \\ (p_i - g_i)x_i - (v_i - g_i)Q_i) & \text{if } Q_i \ge x_i \end{cases} \tag{1.23}$$

and the total profit as $F(x, X(\xi)) = \sum_i^n f(x_i, x_i(\xi_i))$ subjected to a budget constraint. They adopted the credibility theory and defined the credibility of a fuzzy event as $Cr(x_i \ge p) = \frac{1}{2}[Pos(x_i \ge Q_i) + Nec(x_i \ge Q_i)]$, where $Pos(x_i \ge Q_i) = sup_{u \ge Q_i} \mu(u)$ and $Nec(x_i \ge Q_i) = 1 - sup_{u < Q_i} \mu(u)$. The maximum expected profit of the newsvendor problem (MEP) is $E[F(Q^*, x_i)]$ when $E[F(Q^*, x_i)] \ge E[F(Q, x_i)]$ holds for all feasible Q. In the cases where a confidence level, α, is set as a safety margin, α-maximum profit is $\overline{F}$, where $\max(\overline{F}|Cr(F(Q^*, x_i) \ge \overline{F}) \ge \alpha) \ge \max(\overline{F}|Cr((Q, x_i) \ge \overline{F} \ge \alpha)$. The most maximum profit (MMP) is $F(Q^*, x_i)$, when $Cr((Q^*, x_i) \ge F_0) \ge Cr((Q, x_i) \ge F_0)$ where F_0 is the predetermined profit. The authors formulate three models to represent the problem. The first one, the expected value model, maximizes the expected value operator of the fuzzy event with nonnegativity and the budget constraints. The second one, chance constraint model, maximizes $\alpha - MP$ subject to credibility, nonnegativity, and budget constraints. The third model, dependent chance programming, maximizes the credibility not less than the predetermined profit with budget and the nonnegativity constraints. In this paper, a hybrid intelligent algorithm combining fuzzy simulation and genetic algorithm is introduced and numerical examples are provided to display the performance of this algorithm with the three different models mentioned.

Lau and Lau (1988) studied an MPNP, where the objective is to maximize the probability of a given target profit. They assumed that the shortage cost is zero and also showed that any problem with $g_i \ge 0$ can be converted to another one without salvage value. The objective is to maximize $P_T = Prob($Total Profit $\ge$ Target Profit(T) $)$. They consider three different approaches to find the optimum; use simulation to find Q_i^*s and repeat to find the maximizing P_T for different pairs of Q_i^*s, derive an expression for P_T and use a "hill-climbing" method to find Q_i^*s or analytically solve for the FOCs of P_T. They execute approach 2 and 3, and perform some numerical studies for specific demand distributions. In the first case, they define T_m as $(p_1 - v_1)\mu_1 + (p_2 - v_2)\mu_2$, and the target profit as T,which is set to be T_m, $0.5T_m$, or $0.25T_m$. The products are assumed to have identical parameters and demand distributions. Lau proved previously in a single product model that

$$Q_i^* = T/(p - v). \tag{1.24}$$

In order to find the optimum, they first find the Q_I^* that satisfy (1.24) for the single item case. The results show that when $Q_I = Q_i$ the individual and global optimal values are the same. Otherwise, if $P_T^* < P_I$, the optimal for individual products do not give the global optimum. If $P_T^* > P_I$ and $Q_I \neq Q_i$, a reward policy can be implemented to drive the subordinates to achieve the maximum global probability. The authors applied this procedure for normally distributed demands as well. While deriving the mathematical expression for P_T in approach 3, they introduced two different situations. Situation B happens when $p_iQ_i \geq T + v_1Q_1 + v_2Q_2$ for both products and Situation A happens if $p_iQ_i \geq T + v_1Q_1 + v_2Q_2$ holds for one product, and $p_iQ_i < T + v_1Q_1 + v_2Q_2$ holds for the other.

- Situation A

Range 1: $0 \leq x_1 \leq L_1$ where $L_1 = (T + v_1Q_1 + v_2Q_2 - p_2Q_2)/p_1$. We know that the profit from product 2 is $p_2Q2 - v_2Q_2$; thus, product 1 must contribute $T - (p_2Q2 - v_2Q_2)$. If the demand is in this range, $P_{T1} = 0$.

Range 2: $L_1 \leq x_1 \leq Q_1$. In this range, the profit from product 1 is $p_1x_1 - v_1Q_1$ and the profit from product 2 is $T - (p_1x_1 - v_1Q_1)$. Hence, the probability of achieving T when the demand is in range 2 is: $P_{T2} = \int_{L_1}^{Q_1} f_1(x_1)[1 - F_2(\frac{T+v_1Q_1+v_2Q_2-p_1x_1}{p_2})]\mathrm{d}x_1$.

Range 3: $Q_1 \leq x_1 \leq \infty$. In this range, the profit from product 1 is constant, $(p_1 - v_1)Q_1$, and the profit from product 2 is $T - (p_1 - v_1)Q_1$. Thus, the probability of achieving T when the demand is in range 3 is: $P_{T3} = \int_{Q_1}^{\infty} f_1(x_1)[1 - F_2(\frac{T+v_1Q_1+v_2Q_2-p_1Q_1}{p_2})]\mathrm{d}x_1$.

- Situation B

Range 1: $0 \leq x_1 \leq L_2$ where $L_2 = (T + v_1Q_1 + v_2Q_2)p_1$. Hence, $P_{T1} = \int_0^{L_2} f_1(x_1)[1 - F_2(\frac{T+v_1Q_1+v_2Q_2-p_1x_1}{p_2})]\mathrm{d}x_2$.

Range 2: $L_1 \leq x_1 \leq \infty$. In this range, the probability of achieving T is: $P_{T2} = 1 - F_1(L_2)$.

The authors then derive the optimal ordering quantities assuming that the parameters for both items are equal and the demand follows a uniform distribution. Finally, they consider the case where the selling prices of each item is different. They found using approach 2 that if $p_1 < p_2$, then $Q_1^* > Q_2^*$ when T is small and $Q_1^* < Q_2^*$ when $T \geq T_m$.

Vairaktarakis (2000) mentioned in his paper, ".....along with the traditional risk averse attitude, the managers render minimax regret approaches very important in identifying robust solutions, i.e., solutions that perform well for any realization of the uncertain demand parameters." Based on this, he presents a number of minimax regret formulations for the multi-item newsvendor problem with a single budget constraint, when the demand distribution is completely unknown. Demand uncertainty is captured by means of discrete and continuous scenarios.

In discrete demand scenarios, let $D^S(i)$ be the collection of all possible demand realizations for item i, $i = 1, 2, \ldots, n$. Then, the solution to any of the multi-item problems that will be stated below must be n-tuple in $D^S(1) \times D^S(2) \times \ldots \times D^S(n)$. We consider three different objective functions.

- *Absolute robustness*. This approach attempts to find an n-tuple of order quantities that maximize the worst case profit over all possible demand realizations.

$$\max_{(Q_1 \ldots Q_n) \in D^S(1) \times D^S(2) \times \ldots \times D^S(n)} \ \min_{(d_1 \ldots d_n) \in D^S(1) \times D^S(2) \times \ldots \times D^S(n)} \sum_{i=1}^{n} \pi_i(Q_i, d_i)$$

$$s.t. \quad \sum_{i=1}^{n} c_i Q_i \leq W.$$

- *Robust deviation*. This formulation provides a solution that minimizes the maximum profit loss due to demand uncertainty. The objective function is:

$$\min_{(Q_1 \ldots Q_n) \in D^S(1) \times D^S(2) \times \ldots \times D^S(n)} \ \max_{(d_1 \ldots d_n) \in D^S(1) \times D^S(2) \times \ldots \times D^S(n)} \times \sum_{i=1}^{n} \pi_i(d_i, d_i) - \pi_i(Q_i, d_i) \quad s.t. \quad \sum_{i=1}^{n} c_i Q_i \leq W,$$

  where $\pi_i(d_i, d_i) - \pi_i(Q_i, d_i)$ stands for the profit that could be realized if there was no demand uncertainty less the profit made for the order quantity Q_i.
- *Relative robustness*. This minimizes the relative profit loss per unit of profit that could be made if there was no demand uncertainty.

$$\min_{(Q_1 \ldots Q_n) \in D^S(1) \times D^S(2) \times \ldots \times D^S(n)} \ \max_{(d_1 \ldots d_n) \in D^S(1) \times D^S(2) \times \ldots \times D^S(n)} \times \sum_{i=1}^{n} \frac{\pi_i(d_i, d_i) - \pi_i(Q_i, d_i)}{\pi_i(d_i, d_i)} \quad s.t. \quad \sum_{i=1}^{n} c_i Q_i \leq W.$$

In the continuous demand scenario, the demand in (1.4) is bounded by $\underline{D}$ and $\overline{D}$. Then the minmax problem becomes:

$$\max_{Q_i} \ \min_{d_i \in [\underline{D_i}, \overline{D_i}]} \sum_{i=1}^{n} \pi_i(Q_i, d_i)$$

$$s.t. \quad \sum_{i=1}^{n} c_i Q_i \leq W,$$

$$Q_i \in [\underline{D_i}, \overline{D_i}].$$

This problem can be reduced to a continuous knapsack problem, and solved by the proposed algorithm in this paper. Then, the optimal quantity is:

$$Q_* = \frac{(v_i - g_i)\underline{D} + (p_i - v_i + B_i)\overline{D}}{p_i - g_i + B_i}. \tag{1.25}$$

Similarly, Choi et al. (2011) considered a risk-averse MPNP under the law-invariant coherent measures of risk. They have shown that for heterogeneous products with independent demands, increased risk aversion leads to decreased orders, and derived closed-form approximations for the optimal order quantities. Also, they have shown that risk-neutral (maximize the expected profit) solutions are asymptotically optimal under risk aversion as the number of products tends to be infinity. This result has an important business implication: companies with many products or product families with low demand dependence need to look only at risk-neutral solutions even if they are risk averse. For a risk-averse newsvendor with dependent demands, they showed that in a two-product model with positively dependent demands, the optimal order quantities are lower than for independent demands, while for negatively dependent demands, the optimal order quantities are higher.

In another paper where risk was taken into consideration, Ozler et al. (2009) consider a single-period MPNP, where a retailer determines the optimal order quantities of N different products having stochastic demand. Furthermore, they integrate risk considerations (i.e., the risk of earning less than a desired target profit or losing more than an acceptable level due to demand uncertainty) through a Value at Risk (VaR) approach. VaR is a measure of downside risk and is defined as the probability of earning lower than the target profit value is less than or equal to a threshold probability value.

In order to illustrate the approach, the authors first derive a compact expression for the distribution of the profit for two products with a joint demand distribution, and explicitly derive the VaR constraint in terms of the decision variables Q_1 and Q_2. The formulated problem turns out to be a mixed-integer programming formulation with a nonlinear objective function under mixed linear and nonlinear constraints. They analyze the conditions for the feasibility of this problem and present a mathematical programming formulation that determines the optimal order quantities. The authors also consider a correlated demand structure, and by solving the two-product problem, they show that the expected profit is higher when two products with negatively correlated demands are used under a VaR constraint. On the other hand, when the VaR constraint is ignored, demand correlations have no impact on the expected profit.

The authors also attempt to extend the procedure outlined for the two product case to more than two products. In this line, they develop an approximation method in case where there are N products with independent, normally distributed demands. They utilize the central limit theorem to determine the distribution of profit approximately and express VaR constraint by using the normal approximation. Similar to the two-product setting, they analyze the feasibility conditions and present a mathematical programming approach that yields optimal order quantities. The case of the MPNP with a correlated demand structure is left for future research.

Mieghem and Rudi (2002) introduce a class of models called *newsvendor networks*, which allow for multiple products, multiple processing, and storage points and investigate how their single-period properties extend to dynamic settings. Such a model provides a parsimonious framework to study various problems of stochastic capacity investment and inventory management, including assembly, commonality, distribution, flexibility, substitution, and transshipment.

Consistent with the previous multidimensional newsvendor models, the newsvendor networks are defined by a linear production technology, which describes how inputs (supply) is transformed into outputs of fill end–product demand, a linear financial structure, and a probability distribution of end–product demand. This paper continues by incorporating multiple storage points into the multidimensional newsvendor model.

We describe the features of a newsvendor network briefly. Before the demand is realized, a set of "ex-ante" activities are performed on the inputs and their results are stored in "stocks" or inventories. After the demand is realized, "ex-post"activities process stocked inputs into demanded outputs using resources. In addition to being constrained by demand, the sales or the output rate is also constrained both by input stock levels and by the resource capacities, denoted by vectors S and K. The ex-ante activities generate the cost vector, v; the ex-post activities generate the marginal value vector $p-v$; the units carried over to subsequent period incur a holding cost h. Let c_K denote per unit capacity investment cost and x denote the flow units. For example, activities 3 and 2 deplete stocks 1 and 2, respectively, and consume Resource 2's capacity at a rate of α^{-1} and 1; activity 1 depletes stock 1 and consumes resource 1. Hence, the inventory constraints are: $x_1 + x_3 \leq S_1$ and $x_2 \leq S_2$, while the capacity constraints are $x_1 \leq K_1$ and $x_2 + \alpha^{-1} x_3 \leq K_2$. Newsvendor networks are thus about three decisions: capacity investment decisions K, input inventory procurement decisions S, and activity decisions $x(K,S,D)$.

The objective is to maximize the expected operating profit, which is the net value from processing minus the shortage penalty cost and holding cost:

$$\Pi(K,S) = E \max_{x \in X(K,S,D)} \left[(p-v)X - B(D - R_D x)^+ - h(S - R_S x)^+ \right],$$

where R_S and R_D are input–output matrices, and A is the capacity consumption matrix. The set of feasible activities are constrained by supply S, demand D, and capacity K:

$$X(K,S,D) = x \leq 0 : R_S x \leq S, R_D x \leq D, Ax \leq K.$$

The expected firm value to be maximized is :

$$V(K,S) = \Pi(K,S) - vS - c_K K.$$

This paper presented single period optimality conditions and showed that they retain their optimality in a dynamic setting, so that a stationary base-stock policy is optimal. Besides, it also shows that as in most inventory settings, lost sales are more tractable in newsvendor networks than backlogging. The discussion suggests that the culprits are discretionary activities or joint ex-post capacity constraints, both of which make the order-up-to levels of inputs dependent on backlog in a nonlinear manner.

1.2.2 Multiple Constraints

Similar to the MPNP with a single constraint, the MPNP with multiple constraints has also been investigated by a few researchers. Ben-Daya and Raouf (1993) first presented an analytical solution procedure for a two-constraint multi-item newsvendor problem in which all items' demands are uniformly distributed. Lau and Lau (1995) presented a Lagrangian-based numerical solution procedure of a multi-item newsvendor problem with multiple constraints. Their proposed solution procedure is an adaptation of the "Active Set Methods" which consists of two basic components:

- *Component A*. For a given subset W(called the "working set") of all the resource constraints, component A solves the equality-constrained problem:

$$\begin{aligned} &\text{Max} \sum_{i=1}^{N} E[\pi_i(Qi)] \\ &\text{s.t.} \sum_{i=1}^{N} r_{i,j} Q_i \leq R_j, j = 1, 2, \ldots M, \end{aligned} \tag{1.26}$$

  where $r_{i,j}$ is the coefficient of resource j of item i and $R_j$is the amount available of resource j.
- *Component B*. This is the procedure for defining and updating the working set W for each altered component A. This component A and component B cycle is repeated until the optimal condition is met. The authors provide mathematical details and numerical examples to validate this method.

Lau and Lau (1997), in the sequel of their earlier works, proposed a three-step procedure that used subjective probability elicitation to supplement whatever empirical data is available to construct demand distribution functions. Since the typical multi-item newsvendor problem solution procedure requires many repeated evaluations of the demand's inverse cdfs', the authors suggest using Tocher's general "inverse cdf" to fit the distribution function:

$$\begin{aligned} F_T^{-1}(P) &= D \\ &= a + bp + cp^2 + \alpha(1-p)^2 ln(p) + \beta p^2 ln(1-p), \end{aligned} \tag{1.27}$$

where five parameters (a,b,c,α, and β) could be determined by least-squares fitting. Similar to the normal distribution, $F_T^{-1}(P)$ has a negative tail, which is eliminated by using the following modification:

$$F_M^{-1} = D = \max[0, F_T^{-1}(P)]. \tag{1.28}$$

Abdel-Malek and Areeratchakul (2007) and Areeratchakul and Abdel-Malek (2006) developed an approximate solution procedure to deal with this type of problem. It is based on a triangular presentation of the areas resulting from integrals that are included in the objective function, which facilitates expressing the objective function in quadratic terms. One can solve this problem using familiar linear programming packages. The authors have shown that the objective function can be expressed in the following quadratic form:

$$\text{Max}\, Z = \sum_{i=1}^{N} \left(A_i^{(.)} x_i^2 + B_i^{(.)} x_i + C_i^{(.)} \right), \tag{1.29}$$

where expressions $A_i^{(.)}$, $B_i^{(.)}$, and $C_i^{(.)}$ are constants to be determined for each product according to its demand probability distribution. In order to get the quadratic form above, we first need to approximate the integral of the cumulative distribution function using triangular approach as:

$$\int_0^{x_i} F(D_i)\mathrm{d}D_i \approx \frac{1}{2}(x_i - x_{l,i})(\Delta_i(x_i - x_{l,i})), \tag{1.30}$$

where $x_i - x_{l,i}$ is the length of the triangle base, $F(x_i) = \Delta_i(x_i - x_{l,i})$ is the height of the triangle with respect to x_i, and Δ_i represents the slope of the triangle. For more details about these parameters under different distribution functions, readers can refer to the paper.

Abdel-Malek and Montanari (2005b) discussed a solution procedure for the MPNP with two constraints. The methodology in the paper is based on analyzing the dual of the solution space as defined by the constraints of the problem. In order to avoid infeasible (negative) solutions, the authors propose that we begin by defining the possible regions of the dual of the solution space. The corresponding solution method is selected based on the area which the resource point is in. Finally, the authors present two numerical examples to illustrate the application of the proposed approach; the first one considers the case where only one of the constraints is binding, and the second one analyzes the case where both constraints are binding.

In addition to the methods mentioned above, Niederhoff (2007) utilized the separability of the objective function and used convex separable programming to minimize the expected cost and calculate the optimal order quantities. Due to the properties of the piecewise linear approximation method, this problem can be studied without any specific distribution. This method also provides sensitivity analysis which can give us some important insights.

1.2.3 Other Constrained MPNP Approaches

In addition to the classical constrained approaches, several authors focused on the applications of the MPNP to address specific issues. Khouja and Mehrez (1996) formulated a MPNP under a storage or budget constraint such that progressive multiple discounts are offered to sell excess inventory. They provided different algorithms depending on whether the optimal order quantities are large or small. The authors assumed that there is a perfect and positive correlation between the demand at the jth discount price and the demand at the nondiscount price. If the demand for the product at the nondiscount price is high, then discounting the price of the product results in a proportionally high additional demand. Observations on the solution to the constrained problem show that storage (or budget) constraint in an MPNP reduces the service levels (i.e., probability of satisfying demand) and order quantities of all products, when compared to the corresponding levels for the unconstrained problem. Furthermore, the numerical examples that compare multiple and single discount solutions indicate that using multiple discounts instead of discounting just once to the salvage value may result in a different optimal solution.

Shi and Zhang (2010), Shi et al. (2011) and Zhang (2010) investigated the MPNP with supplier quantity discounts and a budget constraint, and the effect of these two features on the optimal order quantities. In this line, Zhang presented a mixed integer nonlinear programming model to formulate the problem. The proposed Lagrangian relaxation approach is demonstrated by means of numerical tests. Finally, the problem is extended to multiple constraints, including space or other resource limitations. It is assumed that suppliers provide all-quantity discounts, and the newsvendor faces uncertain demand for multiple products. Besides, the probability density function for each product is assumed to be given.

To solve the problem, the authors use the Lagrangian heuristic and present methods to find upper and lower bounds, as well as an initial feasible solution. They relax the budget constraint (instead of discount constraints that potentially give a tighter dual bound) as it results in a classical newsvendor subproblem with discount constraints. The computational results indicate that the algorithm is extremely effective for the newsvendor model with supplier quantity discounts and a budget constraint (in terms of both solution quality and computing time). The computational results for the multi-constraint case also indicate that the proposed approach performs well for the problems with multiple constraints.

In a different extension, Chen and Chen (2010) developed a multi-product newsvendor model under a budget constraint with the addition of a reservation policy. Reservation policies reduce the demand uncertainty of newsvendor-type products. Under the reservation policy studied in this paper, a discount rate is offered to consumers in order to induce them to make a reservation and buy in advance. The authors propose a general algorithm, namely the MCR algorithm, which finds the optimal order quantity and the discount rate necessary to maximize the total expected profit under the budget constraint. In order to illustrate the efficiency of the proposed algorithm, MCR, they solve a numerical example and compare the classical multi-product budget-constraint newsvendor model (CMC model) with the multi-product budget-constraint newsvendor model with the reservation policy. Numerical results show that the total expected profit obtained from the MCR is greater than that of CMC. This is tied to the reservation policy proposed in the model. The difference between the profits of these two models is treated as the value of information. Thus, we can conclude that the decision to adopt the reservation policy depends on the trade-off between the information value and the cost incurred to establish the willingness function and extra-demand functions.

Aviv and Federgruen (2001) address the multi-product inventory system problem with random and seasonally fluctuating demands. Moreover, they extend the analysis to a multi-echelon problem, two stages specifically. This paper contributes to the literature on *delayed product differentiation* strategies and makes an assumption that "Demands in each period follow a given multivariate distribution with arbitrary correlations between items." In addition, "....as in virtually all inventory models, the demands in different periods are independent and their distributions are perfectly known." Unsatisfied demand is backlogged; each cost component of a product is a function of the product's inventory position(including inventory on hand, blanks being transformed into units of the final product and minus the backlogs). The objective is to minimize expected discounted cost over a finite or infinite horizon or to minimize the long-run average value.

To include all the cost components, this paper defines a expected value of costs for jth item in a period of type k as

$$\overline{G}_j^k = \alpha^j E h_j \left(\left[y_j - d_j^k - d_j^{k+1} \ldots - d_j^{k+l_j} \right]^+ \right)$$
$$+ p_j \left(\left[d_j^k + d_j^{k+1} + \ldots + d_j^{k+l_j} - y_j \right] + \right),$$

where y_j is the inventory position of item j at the beginning of a period, d_j^k is the demand at period k for item j, h_j is the holding cost, and α^j is the discount factor. The model can be formulated as a Markov Decision Process with countable state space $S = \{(x,k)$, x is integer, $k = 1,...,K\}$ and finite action sets $A(x,k) = \{y : x \leq y$ and $\sum_{j=1}^{l} y_j \leq \sum_{j=1}^{l} x_j + b^k\}$. To solve this problem, the authors propose a lower-bond approximation and heuristic strategies. In the case of a 2-stage echelon, i.e., production has positive lead time, they simply modified $R^k(.)$, which is the one-step cost function in a single-item model, and introduced the system-wide echelon inventory position of blanks.

Chung et al. (2008) considered the items with short life cycles or seasonal demands. They developed a two-stage, multi-item model incorporating the reactive production that employs a firm's internal capacity. Reactive capacities are preallocated to each item in preseason stage and cannot be changed during the reactive stage. The objective is expected profit maximization. A simple algorithm for computing optimal policies is presented. This paper aims to help managers understand how employing internal capacity during reactive stage can reduce the impact of the poor demand forecasts. Without fixed costs, the optimal production vector for the reactive stage is a simple function of the production vector, Q, for the preseason stage and the capacity allocation vector, Z, for the reactive stage. By analyzing the KKT conditions, the optimal solution and the Lagrange multiplier, λ^*, can be determined.

Casimir (2002) used MPNP to determine the value of incomplete information. He focused on the value of information in three newsvendor models: the basic model with no constraints, the model with budget or capacity constraints, and the model with substitutability. The value of incomplete information is considered in the form of product-mix information and global information. Product-mix information implies total demand is unknown, but the distribution over products is known exactly. In this case, the overall optimal order quantity is determined initially, and then the optimal order quantity for each single item is determined from the actual value. In the case of global information, total demand is known, but its distribution over the products is unknown. Then, the optimal order quantity for each individual item with the given total demand has to be determined. Here, the authors compute the value of incomplete information by comparing the expected profits of the two cases, and do not consider the performance criteria. Besides, rather than the computation of optimal order quantities, results are computed numerically to provide a research framework for the value of information. Their assumptions are: (1) demand is normally distributed, (2) demand for different items is independent, (3) the salvage value for unsold items is zero, (4) there is no penalty for unmet demand, (5) price, cost, average demand, and standard deviation of demand for all products are the same, (6) for the model with budget constraint, only two items are considered, (7) analyzing substitutability, only a two-item newsvendor problem is considered, and it is assumed that the customer takes a single unit of the substitution product, and (8) substitutability is assumed to be symmetric.

In the model without additional complications, it is shown that the value of product-mix information increases with the number of items, whereas the value of global information decreases with the number of items. The value of both product-mix information and global information decreases with a budget constraint. Furthermore, the value of perfect information also decreases with a budget constraint. The probability of substitution decreases the value of product-mix information such that it is zero with complete substitution, and increases the value of global information so that it is equal to the value of perfect information with complete substitution.

Finally, Zhang and Hua (2010) consider a system where the products are procured from the supplier with a fixed-price contract. Under this procurement strategy, the retailer does not order enough products to avoid the risk raised from demand uncertainty (i.e., lower realized demand). The authors here apply a portfolio approach to MPNP under a budget constraint, where the retailer's procurement strategy is designed as a portfolio contract. In this case, each newsvendor product can be procured from the supplier with dual contracts: a fixed-price contract and an option contract. The retailer can lower the inventory risk by utilizing the flexibility of the option contract. On the other hand, it in turn results in additional costs compared to fixed-price contracts, since unit reservation and execution cost of option contract is typically higher than unit cost of a fixed-price contract. In the paper, the objective is to maximize the total expected profit of the retailer through determining the optimal order quantities of products procured with portfolio contracts. The authors consider a single-period model and assume that the retailer sells n products with independent and stochastic demands. All demands are considered to be nonnegative. The portfolio contract consists of a fixed-price contract and an option contract. In the fixed-price contract, the retailer pays a unit fixed cost for each product he procured from the supplier. In the option contract, to reserve certain order quantity, the retailer prepays a unit reservation cost up-front. Then, the retailer pays an execution cost for each unit purchased up to the option reservation level. The retailer loses the initial payment if he does not exercise the option. Related to those, it is further assumed that:

- The total cost of option contract (reservation plus execution cost) is larger than the cost of fixed-price contract.
- The reservation cost of option contract is smaller than the pure procurement cost of the fixed-price contract.

Following the problem formulation, the authors establish the structural properties of the optimal solution (e.g., the concavity of expected profit function) and propose a polynomial solution algorithm of o(n) order. The main advantage of the proposed algorithm is that it does not depend on a specific demand distribution and it is applicable to general continuous demand distributions. Finally, they conduct numerical studies and sensitivity analysis to show the efficiency of the proposed algorithm, as well as compare three procurement contracts: fixed-price contract (FC), option contract (OC), and portfolio contract (PC). It is evident that the newsvendor model with PC, generates significant improvement compared to FC and OC models. Furthermore, it is shown that following the increase in the available budget, the performance gap between FC and PC models decreases, while the gap between OC and PC increases.

Vaagen and Wallace (2008) study risk hedging in fashion supply chains. They consider two states of the world: $State1$ when a variant of a product becomes popular and the others go out of fashion, and $State2$ is when the reverse happens. In this paper, Vaagen and Wallace provide a portfolio building decision model under uncertainty by combining The Markowitz and the newsboy models into a stochastic optimization model. This model tries to minimize the profit risk using semi-variance. The results of the different scenarios show that hedging portfolios gives any company a competitive advantage. We can also conclude that the uncertain information such as demand estimates and trend information for a certain group of products are not as important in the fashion industry as it is in other industries. The best approach in this case is to define and release hedge portfolios. This model can be extended to include substitutability, which is discussed by Cheng and Choi (2010).

1.3 Substitute Products

Retailers often offer product substitutes to prevent customer loss. This substitution can be perfect, partial, or downward. Most of the early works used two-way substitution and introduced heuristics to find the optimal order quantities. Recent works, however, focus more on one-way substitution. This type of substitution arises in real life, for example, in the semiconductor industry; a higher capacity chip can be used to satisfy demands for lower capacity chips. Current literature in this stream can be classified as those focusing on one-way substitution or two-way substitution.

1.3.1 One-Way Substitution

Bassok et al. (1999) concentrated on full downward substitution among the various structures of substitution. Considering that there are N products and N demand classes, full downward substitution implies that excess demand for class i can be satisfied using stocks of product j for $i \geq j$. The authors discuss a two-stage profit maximization formulation for the multi-product substitution problem. In the first stage, the orders are placed (before demands are realized) , and in the second stage the products are allocated to demands (after demand is observed) (i.e., allocation problem). The authors assume that there are N products and N demand classes, and the demands for each class are stochastic. The order, holding, penalty, and salvage costs are assumed to be proportional, and the revenue is linear in the quantity sold. It is further assumed that the substitution cost is proportional to the quantity substituted. Delivery lags and capacity constraints are ruled out. Finally, it is assumed that the revenue earned for each unit of satisfied demand in class i depends only on i and not on the type of product j used to satisfy the excess demand. The authors assume that: (a) it is more profitable to satisfy unmet demand of class i than of class j, for $i < j$; (b) the effective salvage value of product i is not less than that of product j, for $i < j$; and (c) the substitution of product i for demand class j is profitable.

Let $I(\overrightarrow{x})$ be the maximum single period profits and $P(\overrightarrow{x},\overrightarrow{y})$ be the expected single period profits when the starting inventory before placing the order is $\overrightarrow{x}$ and after ordering is raised to $\overrightarrow{y}$. Then, $I(\overrightarrow{x}) = \max_{(}\overrightarrow{y} \geq \overrightarrow{x})P(\overrightarrow{x},\overrightarrow{y})$. Let $\overrightarrow{d} = (d_1,\ldots,d_N)$ be a vector of realized demands. Define $F(\overrightarrow{d}) = F_{1,2,\ldots,N}(d_1,\ldots,d_N)$ as the joint distribution of demands from class 1 to N. Let $G(\overrightarrow{y},\overrightarrow{d})$ be the profits for a given stock level, $\overrightarrow{y}$, and the realized demand, $\overrightarrow{d}$. Let $w_i j$ be the quantity of product j allocated to the demand class j. Then

$$P(\overrightarrow{x},\overrightarrow{y}) = -\sum_{k=1}^{N} c_k(y_k - x_k) + \int_{R_+^N} G(\overrightarrow{y},\overrightarrow{d})\mathrm{d}F(\overrightarrow{d}), \tag{1.31}$$

where:

$$G(\overrightarrow{y},\overrightarrow{d}) = \max_{u_i,v_i,w_{ji}} \sum_{i=1}^{N}\sum_{j=1}^{i} a_{ji}w_{ji} + \sum_{i=1}^{N} s_i v_i - \sum_{i=1}^{N} \pi_i u_i.$$

Subject to

$$u_i + \sum_{j=1}^{i} w_{ji} = d_i \text{ for } i = 1,\ldots,N, v_i + \sum_{i=1}^{N} w_{ji} = y_i \text{ for } i = 1,\ldots,N. \tag{1.32}$$

The authors present a greedy algorithm for the allocation problem, and give a new and compact notation of writing the first differentials of the profit function with respect to stock levels. They prove that given a starting inventory level, the allocation algorithm will maximize profits in $G(\overrightarrow{y}, \overrightarrow{d})$. In addition, the profit function $P(\overrightarrow{x}, \overrightarrow{y})$ is proven to be concave and submodular. They also propose an iterative algorithm to compute the order points for a two-product problem, and develop bounds on the optimal order points. Finally, they present a computational study for the two-product problem and show that the benefits of solving for the optimal quantities, when substitution is considered at the ordering stage, are higher with high demand variability, low substitution cost, low profit margins, high salvage values, and similarity of products in terms of prices and costs.

Smith and Agrawal (2000) have analyzed the impact of retail assortments on inventory management and customer service. They focused on product variety in retailing environment, where the customers can often be satisfied by one of several items, e.g., light colors of T-shirts in apparel. In this paper, they develop a probabilistic demand model for items in an assortment that capture the effects of substitution and provide a methodology for selecting item inventory levels so as to maximize total expected profit, subject to given resource constraints. Because of substitution, the inventory levels for products in an assortment must be optimized jointly.

They consider inventory policies that reinitialize at the start of each fixed cycle, assuming lost sales occur if there is a stockout before the end of the cycle. The authors also analyze several illustrative numerical examples to demonstrate the insights, such as, substitution effects can reduce the optimal assortment size, and policies of ignoring substitution effects can be less profitable than those that explicitly incorporate substitution effects. Similarly, Shah and Avittathur (2007) studied the effects of retail assortments on inventory control. Different from Smith,he defined a demand cannibalization and substitution index and assumed the demand to be a Poisson process (similar to Anupindi et al. 1998). The numerical results showed that when the fixed cost and salvage value of a customized product is high and its incremental profits are low, it is not feasible to carry customized products.

In addition, Smith and Agrawal (2000) also studied a "static" substitution model. They assumed that the choice by the customer is independent of the current inventory levels and the customer does not accept a second choice. Mahajan and van Ryzin (2001) used a choice process based on a utility that is assigned by the customer to each product. This utility is interpreted as the net benefit to the customer from purchasing or not purchasing a product. In this case, the information available to the retailer is only the probability of a sample path $\omega = \{U_t : t + 1, \ldots, T\}$, where T is the number of customers. The number of sales is dependent on the initial inventory level and the sample path. The authors then introduce a sample path gradient algorithm to obtain the optimal results.

Dutta and Chakraborty (2010) studied the newsboy problem with one-way substitution where the demand is fuzzy. The membership function of demand of product i is represented as:

$$\mu_{\widetilde{D_i}(x)} = \begin{cases} L_i(x) = \frac{x-\underline{D_i}}{D_i-\underline{D_i}}, & \underline{D_i} \le x \le D_i \\ R_i(x) = \frac{\overline{D_i}-x}{\overline{D_i}-D_i}, & D_i \ge x \ge \overline{D_i}, \\ 0, & \text{otherwise}, \end{cases} \tag{1.33}$$

where the demand is $\widetilde{D_i} = (\underline{D_i}, D_i, \overline{D_i})$. The fuzzy objective function is complex and the concavity proof is difficult; therefore, Dutta and Chakraborty developed an algorithm to find the optimal order quantity. They defined four situations of demand in relation to the Q^* and run the complete procedure for each one of these. They ran some numerical examples to provide validation for their method and made recommendations for further research to include salvage value and holding cost as well as two-way substitution.

Considering a stylized scenario for two products, without a loss of generality, assume that product 1 substitutes for product 2 one-to-one, and if there is a substitution, this item is sold at the price of product 2. We also assume that the selling price of the substituted item is higher than the cost of the substitute as well as its salvage value. Then the actual end of period profit for the buyer is:

$$\begin{array}{ll} \textit{Case 1.}\ \sum_{i=1}^{2}[p_i x_i - v_i Q_i + g_i(Q_i - x_i)] & \text{if } x_1 \le Q_1; x_2 \le Q_2 \\ \textit{Case 2.}\ p_1Q_1 + p_2x_2 - \sum_{i=1}^{2} v_iQ_i + g_2(Q_2 - x_2) - B_1(x_1 - Q_1) & \text{if } x_1 > Q1; x_2 \le Q_2 \\ \textit{Case 3.}\ p_1x_1 + p_2Q_2 - \sum_{i=1}^{2} v_iQ_i + p_2 Min(x2 - Q_2, Q_1 - x_1) & \text{if } x_1 \le Q1; x_2 > Q_2 \\ \quad +g_1[Q_1 - x_1 - (x_2 - Q_2)]^+ - B_2[x_2 - Q_2 - (Q_1 - x_1)]^+ & \\ \textit{Case 4.}\ \sum_{i=1}^{2} p_iQ_i - v_iQ_i - B_i(x_i - Q_i) & \text{if } x_1 > Q1; x_2 > Q_2 \end{array} \tag{1.34}$$

and based on this, the expected profit function is:

$$\begin{aligned} E[\pi(Q_1,Q_2)] = E\Bigg[& [p_1 Min(x_1,Q_1) + p_2 Min[x_2, Q_2 + (Q_1 - x_1)^+] - v_1Q_1 - v_2Q_2 \\ & + g_1[Q_1 - x_1 - (x_2 - Q_2)^+]^+ + g_2(Q_2 - x_2)^+ - B_1(x_1 - Q_1)^+ \\ & - B_2[x_2 - Q_2 - (Q_1 - x_1)^+]^+ \Bigg] \\ = p_1 & \left[\int_0^{Q_1} x_1 f_1(x_1) dx_1 + \int_{Q_1}^{\infty} Q_1 f_1(x_1) \mathrm{d}x_1\right] \\ & + p_2 \left[\int_0^{Q_2} x_2 f_2(x_2) \mathrm{d}x_2 + \int_{Q_2}^{\infty} Q_2 f_2(x_2) \mathrm{d}x_2\right. \end{aligned}$$

$$
\begin{aligned}
&+ \int_0^{Q_1}\int_{Q_2}^{\infty}(Q_1-x_1)f(x_1,x_2)\mathrm{d}x_2\mathrm{d}x_1\Bigg]\\
&- v_1Q_1 - v_2Q_2 + g_1\left[\int_0^{Q_1}\int_0^{Q_2}(Q_1-x_1)f(x_1,x_2)\mathrm{d}x_2\mathrm{d}x_1\right.\\
&\left.+\int_0^{Q_1+Q_2-x_2}\int_{Q_2}^{Q_1+Q_2}(Q_1+Q_2-x_1-x_2)f(x_1,x_2)\mathrm{d}x_2\mathrm{d}x_1\right]\\
&+ g_2\int_0^{Q_2}(Q_2-x_2)f_2(x_2)\mathrm{d}x_2 - B_1\int_{Q_1}^{\infty}(x_1-Q_1)f_1(x_1)\mathrm{d}x_1\\
&- B_2\left[\int_{Q_1}^{\infty}\int_{Q_2}^{\infty}(x_2-Q_2)f_1(x_1,x_2)\mathrm{d}x_2\mathrm{d}x_1\right.\\
&\left.+\int_0^{Q_1}\int_{Q_1+Q_2-x_1}^{\infty}(x_1+x_2-Q_1-Q_2)f(x_1,x_2)\mathrm{d}x_2\mathrm{d}x_1\right]. \quad (1.35)
\end{aligned}
$$

Cai et al. (2004) used a similar expected profit function as above and proved that it is concave and submodular. Using this property, the optimal order quantities can be found by setting the derivatives with respect to Q_1 and Q_2 equal to zero. If we define $G(Q_1,Q_2) = \int_0^{Q_1}\int_0^{Q_1+Q_2-x_1} f(x_1,x_2)\mathrm{d}x_2\mathrm{d}x_1$, the following holds:

$$
F_1(Q_1^*) + \frac{(p_2+B_2)-g_1)}{(p_1+B_1)-(p_2+B_2)}G(Q_1^*,Q_2^*) = \frac{(p_1+B_1)-v_1}{(p_1+B_1)-(p_2+B_2)}, \quad (1.36)
$$

$$
F_2(Q_2^*) + \frac{(p_2+B_2)-g_1)}{(p_2+B_2)-g_2)}\left[G(Q_1^*,Q_2^*) - F(Q_1^*,Q_2^*)\right] = \frac{(p_2+B_2)-v_1}{(p_2+B_2)-g_2}. \quad (1.37)
$$

$F_i(Q_i^*)$ represents the probability of all of the demand for item i being satisfied when the stock level is Q_i^*. $G(Q_1^*,Q_2^*)$ is the probability that the total demand is satisfied given that item 1 was substituted for item 2. $F(Q_1^*,Q_2^*) = \int_0^{Q_1}\int_0^{Q_2} f(x_1,x_2)\mathrm{d}x_2\mathrm{d}x_1$ is defined as the probability that the demand for each item is satisfied without any substitution. Finally, $F_2(Q_2^*)+G(Q_1^*,Q_2^*)-F(Q_1^*,Q_2^*)$ is the probability that all of the demand for item 2 is satisfied using either of the items. Cai et al proved four different properties of the optimal order quantities. Property 1 shows that as the unit price of item i increases, Q_1^* decreases and Q_2^* increases and, evidently Q_2^* decreases as the unit price of item 2 increases. Property 2 states that when the price of each item increases, their respective optimal quantities decrease. Conversely, the increase in price of item 1 decreases the optimal quantity for item 2. Property 3 states a similar argument related to salvage cost. Property 4 indicates that the optimal order quantity of each item is linearly related to their respective mean demands. Property 5 states that the variance of item i affects the optimal quantity of item j reversely. In this paper, the authors showed that the expected profits and the fill rate can be improved by using substitution.

Table 1.1 Notations for two-way substitution

Symbol	Meaning
R	Review period
L_i	Replenishment lead time
$L_i + R_i$	Replenishment cycle
$f_{x^i}(x_0)$	Density function of demand over the replenishment cycle for product i
β_i	Parameter that satisfies, $\sigma_i = \beta_i \sigma_1$
r	Inventory holding cost
S_i	Order up-to level
K_i	Safety factor, $S_i = \overline{x_i} + K_i \sigma_i$
$0 < \alpha_{ij} < 1$	Probability that a costumer will substitute j for a unit of i
$f_u(u_0)$	Density function of standard normal distribution
$G_u(k)$	Tabulated function of the standard normal distribution

1.3.2 Two-Way Substitution

Unlike the one-way substitution, in two-way substitution case, each of the items can be used to supply the demand for another one. This only occurs when the demand for one item is higher than the quantity ordered and the demand for the substitute item is lower than the quantity ordered. McGillivray and Silver (1978) and Parlar and Goyal (1984) assumed whenever substitution is possible, there is a probability that a customer will accept a substitute product. In Parlar's case, this probability was between 0 and 1, whereas it was fixed for McGillivray. In McGillivray's paper, the demand, x_i , is assumed to be normally distributed with a mean of $\bar{x}_i$ and a standard deviation of $\sigma_i = \beta_i \sigma 1$. The order up to level is given as $S_i = \bar{x}_i + K_i \sigma_i$ and the expected shortage per replenishment cycle is $ESPRC_i = \sigma_i G_u(K_i)$. The unit variable costs and shortage cost of the substitutable, items are also assumed to be identical. This assumption is justified by the fact that in reality when two items are substitutable they will have similar prices. Different levels of substitutability were considered in the paper. The notation used in their paper is shown on Table 1.1.

We know that $G_u(k) = \int_k^\infty (u_0 - k) f_u(u_0) \mathrm{d}u_0$, and $\frac{\mathrm{d}G_u(K_i)}{\mathrm{d}K_i} = -P_{u \geq}(K_i)$. By setting the partial derivative of ETRC with respect to K_i to 0, we find that $P_{u \geq}(K_i *) = \frac{Rvr}{B}$ for $i = 1, \ldots, N$. Using the standard normal property $f_u(u_0) = P_{u \geq}(u_0) + G_u(u_0)$, ETRC can be reduced to:

$$ETRC(K_1^*, K_2^*, .., K_n^*) = \frac{1}{2} DR^2 vr + \sigma_1 B f_u(K^*) \sum_{i=1}^{N} \beta_i. \tag{1.38}$$

If we assume that there is full demand transferability and all items are perfect substitutes of each other, $a_{ij} = 1$, a shortage happens only when the total demand for all items is smaller than the total stock up-to level. The total shortage and on-hand inventory decrease the same amount by the transfer sales; therefore, the total net stock stays the same as the general case. Consequently, the expected total relevant costs with perfect substitution is:

$$ETRC_t(K_{t1}^*, K_{t2}^*, .., K_{tn}^*) = \frac{1}{2}DR^2vr + \sigma_1 B\sqrt{\sum \beta_i^2} f_u(K^*). \tag{1.39}$$

The minimum cost equation is the same as the single item newsvendor model with the demand equivalent to the total demand of substitutable item problem. The savings from the substitutability can be expressed as $ETRC(K_1^*, K_2^*, .., K_n^*) - ETRC_t(K_{t1}^*, K_{t2}^*, .., K_{tn}^*)$. Given (1.38) and (1.39) , the maximum possible savings will occur when substitutability, a_{ij}, is equal to 1. Thus,

$$MPS = \sigma_1 B\left(\sum \beta_i - \sqrt{\sum \beta_i^2}\right) f_u(K^*)), \text{ and } K_1^* = ... = K_N^* = K^* = \frac{\sum \beta_i K_{ti}^*}{\sqrt{\sum \beta_i^2}}. \tag{1.40}$$

In addition to these results, the authors (through a numerical analysis for a two item model) show that when both items are substitutable to each other, the potential savings increase when K^* increases. In the case of one way substitution, where $a_{12} = 0$ and $a_{21} = 1$, the optimal policy is to stock item 1 only. This theorem also holds when $0 < a_{12} < 1$ and $a_{21} = 1$. There is no analytic expression for $ESPRC$ when both items are partially substitutable to each other. Hence, McGillivray and Silver simulated a two-item inventory problem with substitutability. As a result, they demonstrated that when substitutability levels are between 0 and 0.75, the model acts as an independent item inventory control problem. Furthermore, a cost penalty larger than 20% of the MPS only occurs when one of the items is a perfect substitute of the other. For the case of partial substitutability, a heuristic approach was developed and tested.

In relation to the heuristic approach, the expected transferred demands were defined as $E(T_{21}) = a_{21}ESPRC_2$ and $E(T_{12}) = a_{12}ESPRC_1$ for items 1 and 2, respectively. This approach tries to find the optimal values for K and S using $P_{u\geq}(K_i^*) = Rvr/[B(1-a_{ji}) + a_{ji}Rvr]$ and $S_i = [\bar{x}_i + a_{ji}\sigma_j G_u(K_j)] + K_i\sigma_i$ for $i \neq j$ $i = 1, \ldots, N$. This two-item model can also be extended to include multiple items and it is computationally straightforward.

Netessine and Rudi's paper Netessine and Rudi (2003) examines the optimal inventory stocking policies for a given product line under the notion that consumers who do not find their first-choice product in the current inventory might substitute a similar product for it (consumer-driven substitution). Namely, there is an arbitrary number of products and each consumer has a first choice product. If this product is out of stock, the consumer might choose one of the other products as a substitute.

Let α_{ij} denote the probability that a customer will substitute j for a unit of i. The demand vector, $D = (D_1, ..., D_n)$, follows a known continuous multivariate demand distribution with positive support. In the centralized inventory model, the expected profit of the company who manages n products is:

$$\pi = E\sum_i \left[p_i \min\left(D_i + \sum_{j\neq i} \alpha_{ij}(D_j - Q_j)^+, Q_i \right) \right.$$
$$\left. -v_i Q_i + g_i \left(Q_i - \left(D_i + \sum_{j\neq i} \alpha_{ij}(D_i - Q_j)^+ \right) \right)^+ \right].$$

The demand vector for i is $D_i^S = D_i + \sum_{j\neq i} \alpha_{ij}(D_j - Q_j)^+$, where the superscript S indicates that the effect from substitution has been accounted for. In other words, D_i^S is the sum of the first-choice demand and demand from substitution. It is conventional to define $u_i = p_i - v_i$, the unit underage cost; and $o_i = v_i - g_i$,the unit overage cost. This paper proves that the first-order necessary optimality conditions of the centralized problem are given by:

$$Pr(D_i < Q_i^c) - Pr(D_i < Q_i^C < D_i^S)$$
$$+ \sum_{j\neq i} \frac{u_i + o_j}{u_i + o_i} \alpha_{ij} Pr(D_j^S < Q_i^C, D_i > Q_i^C) = \frac{u_i}{u_i + o_i}.$$

In the decentralized inventory model, the profit for each firm i is:

$$\pi_i = E\left[u_i D_i^S - u_i (D_i^S - Q_i)^+ - o_i \left(Q_i - D_i^S \right) \right], i = 1, \ldots n.$$

This paper also shows that any Nash equilibrium is characterized by the following optimality conditions:

$$Pr(D_i < Q_i^d) - Pr(D_i < D_i^d < D_i^S) = \frac{u_i}{u_i + o_i}.$$

After comparing the optimal ordering quantity Q_i^c and Q_i^d, the paper finds that: there exist situations when $Q_i^c \geq Q_i^d$ for some i, it is always true that $Q_i^c \leq Q_i^d$ for at least one i, suppose that all the costs are independent and identically distributed, and the consumers are equally likely to switch to any of the $(N-1)$ products for all i, j. Then $Q_i^c \leq Q_i^d$ for all i.

Nagarajan and Rajagopalan (2008) took a different approach and assumed the demands of products to be correlated. They defined the total demand to be D, and the demand portions of the products to be p, $(1-p)$. Without loss of generality, D is set to be 1, and the optimal order quantities for product 1 and 2 differ from the general newsvendor solution by $(1-\gamma)$. This indicates that the higher the fraction of substitution, the lower the inventory levels. In the case of asymmetric costs and random total demand, a fixed proportion, γ_i, of the customers looking for item i when it is depleted will purchase the substitute and $(1-\gamma_i)$ of them will not make a purchase. If we let $\gamma_i^* = \max\{(p_i + B_i) - (h_i + 2c_i)/(p_j + h_j + B_i), 1\}, i, j = 1, 2, i \neq j$ then, if $\gamma_i \leq \gamma_i^*$, the "partially decoupled" inventory policy is optimal. It is evident that when product 2 is priced higher, the optimal base stock for product 1 is lower. Especially, in the case of high enough p_2 and h_2, this base stock level can be below the mean or even close to zero. This means that the risk-pooling effect of substitution reduces the inventories of both products. This effect is more apparent for the inventory of the lower priced product. The authors show that this method can easily be applied to the n-product and multi-period model.

Focusing on a stylized two-product setting, the end of period profit for the buyer are:

$$\begin{aligned}
&\textit{Case 1.}\ \sum_{i=1}^{2}[p_i x_i - v_i Q_i + g_i(Q_i - x_i) && \text{if } x_1 \le Q_1; x_2 \le Q_2\\
&\textit{Case 2.}\ p_1 Q_1 + p_2 x_2 - \sum_{i=1}^{2} v_i Q_i + p_1 \mathrm{Min}(x_1 - Q_1, Q_2 - x_2) && \text{if } x_1 > Q1; x_2 \le Q_2\\
&\quad + g_2[(Q_2 - x_2) - (x_1 - Q_1)]^+ - B_1[(x_1 - Q_1) - (Q_2 - x_2)]^+ &&\\
&\textit{Case 3.}\ p_1 x_1 + p_2 Q_2 - \sum_{i=1}^{2} v_i Q_i + p_2 \mathrm{Min}(x2 - Q_2, Q_1 - x_1) && \text{if } x_1 \le Q1; x_2 > Q_2\\
&\quad + g_1[(Q_1 - x_1) - (x_2 - Q_2)]^+ - B_2[(x_2 - Q_2) - (Q_1 - x_1)]^+ &&\\
&\textit{Case 4.}\ \sum_{i=1}^{2} p_i Q_i - v_i Q_i - B_i(x_i - Q_i) && \text{if } x_1 > Q1; x_2 > Q_2
\end{aligned} \tag{1.41}$$

and based on this, the expected profit function is:

$$\begin{aligned}
E[\pi(Q_1,Q_2)] = E\Bigg[&[p_1 Min(x_1,Q_1) + p_2 Min[x_2, Q_2 + (Q_1 - x_1)^+] - v_1 Q_1 - v_2 Q_2\\
&+ g_1[Q_1 - x_1 - (x_2 - Q_2)^+]^+ + g_2(Q_2 - x_2)^+ - B_1(x_1 - Q_1)^+\\
&- B_2[x_2 - Q_2 - (Q_1 - x_1)^+]^+\Bigg]\\
= p_1 &\left[\int_0^{Q_1} x_1 f_1(x_1)\mathrm{d}x_1 + \int_{Q_1}^{\infty} Q_1 f_1(x_1)\mathrm{d}x_1\right.\\
&\left. + \int_0^{Q_1}\int_{Q_2}^{\infty}(Q_2 - x_2) f(x_1,x_2)\mathrm{d}x_2\mathrm{d}x_1\right]\\
+ p_2 &\left[\int_0^{Q_2} x_2 f_2(x_2)\mathrm{d}x_2 + \int_{Q_2}^{\infty} Q_2 f_2(x_2)\mathrm{d}x_2\right.\\
&\left. + \int_0^{Q_1}\int_{Q_2}^{\infty}(Q_1 - x_1) f(x_1,x_2)\mathrm{d}x_2\mathrm{d}x_1\right]\\
- v_1 Q_1 - v_2 Q_2 + g_1 &\left[\int_0^{Q_1}\int_0^{Q_2}(Q_1 - x_1) f(x_1,x_2)\mathrm{d}x_1\mathrm{d}x_2\right.\\
&\left. + \int_0^{Q_1+Q_2-x_2}\int_{Q_2}^{Q_1+Q_2}(Q_1 + Q_2 - x_1 - x_2) f(x_1,x_2)\mathrm{d}x_1\mathrm{d}x_2\right]\\
+ g_2 &\left[\int_0^{Q_1}\int_0^{Q_2}(Q_2 - x_2) f(x_1,x_2)\mathrm{d}x_2\mathrm{d}x_1\right.
\end{aligned}$$

$$
\begin{aligned}
&+ \int_0^{Q_1+Q_2-x_1} \int_{Q_1}^{Q_1+Q_2} (Q_1+Q_2-x_1-x_2) f(x_1,x_2) \mathrm{d}x_2 \mathrm{d}x_1 \Bigg] \\
&- B_1 \Bigg[\int_{Q_1}^{\infty} \int_{Q_2}^{\infty} (x_1-Q_1) f(x_1,x_2) \mathrm{d}x_1 \mathrm{d}x_2 \\
&\quad + \int_{Q_1+Q_2-x_2}^{\infty} \int_0^{Q_2} (x_1+x_2-Q_1-Q_2) f(x_1,x_2) \mathrm{d}x_2 \mathrm{d}x_1 \Bigg] \\
&- B_2 \Bigg[\int_{Q_1}^{\infty} \int_{Q_2}^{\infty} (x_2-Q_2) f(x_1,x_2) \mathrm{d}x_2 \mathrm{d}x_1 \\
&\quad + \int_0^{Q_1} \int_{Q_1+Q_2-x_1}^{\infty} (x_1+x_2-Q_1-Q_2) f(x_1,x_2) \mathrm{d}x_2 \mathrm{d}x_1 \Bigg].
\end{aligned} \tag{1.42}
$$

Pasternack and Drezner (1991) proved that this function is concave and showed that the optimal quantities can be found using a specific distribution and parameters. Assuming that the revenue from substitution is different from the revenue from regular sales, the authors also explored the effect of substitution on the order quantities and showed that for a given revenue of t_2 for each product 2 that is substituted for product 1:

$$\frac{\mathrm{d}Q_1^*}{\mathrm{d}t_1} > 0 \text{ and } \frac{\mathrm{d}Q_2^*}{\mathrm{d}t_1} < 0, \tag{1.43}$$

and a similar result holds for product 1. This result implies that if the revenue from substitution of one product increases, the optimal order quantity for the other product will decrease and the substitute product will increase. The authors analytically solved the case for the one-way substitution and reached similar insights. In addition, the authors explored the effect of substitution on the total inventory levels and observed that when the revenue from substitution increases, the optimal quantity of the substitutable product increases faster than the substitute.

Rajaram and Tang (2001) studied the same problem but allowed the substitution parameter to be anywhere between 0 and 1. The heuristic they presented explores how the demand variation and correlation as well as the substitution affect the expected order quantities and expected profits. Khouja et al. (1996) used Monte Carlo simulation to find the optimal order quantities. Six events are defined to represent this model. First event is when the demand for each item is less than its order quantities. Second event is when the demand for each item is equal to or higher than its order quantities. The third and the fourth events are when the demand for item 1 is greater than the order quantity, and the excess quantity of item j is sufficient or insufficient, respectively. Similar case holds for the fifth and the sixth events. They define the upper and lower quantity bounds for each item and prove that the optimal quantities will be between these two values. The first property, which aids the proof of Lemma 1, states that it is more profitable to sell customers one unit of i than to sell t_j quantity of item j. They define the lower bounds to be Q_1^{L} and Q_2^{L}, where $F_1(X_1 = Q_1^{\mathrm{L}}) \approx 1$ and $F_2(X_2 = Q_2^{\mathrm{L}}) \approx 1$ holds. Lemma 1 indicates that the optimal solution will always be higher than the lower bound. In order to prove this, three scenarios that violate lemma 1 are considered. They show that for each of the cases, the expected profit increases when the solution is equal to or higher than the lower bound. They define upper bounds to be Q_1^U and Q_2^U, where $F_3(X_3 = Q_1^U) \approx 1$ and $F_4(X_4 = Q_2^U) \approx 1$, $X_3 = X_1 + t_1(X_2 - Q_2^L)$ and $X_4 = X_2 + t_2(X_1 - Q_1^L)$. Using similar arguments to lemma 1, they prove that the optimal solution is lower than the upper bound. Stricter upper bounds can be found assuming X_1 and X_2 are normally distributed; consequently, X_3 and X_4 can be assumed to be normally distributed. They prove that any optimal solution will be less than Q_1^N and Q_2^N , where Q_1^N and Q_2^N are the solutions to the newsvendor problems with demands X_3 and X_4. Numerical tests were run to gain insights to the problem. As a result, it was found that as t_1 increases, Q_1 increases and Q_2 decreases. This can be explained by the decrease in the effective cost of underestimating item 2. Consequently, the demand for item 1 increases and the demand for item 2 decreases.

The assumptions for this paper are: (1) demand is normally distributed, (2) demand for different items is independent, (3) the salvage value for unsold items is zero, (4) there is no penalty for unmet demand, (5) price, cost, average demand, and standard deviation of demand for all products are the same, (6) for the model with budget constraint, only two items are considered, (7) analyzing substitutability, only a two-item newsvendor problem is considered, and it is assumed that the customer takes a single unit of the substitution product, and (8) substitutability is assumed to be symmetric.

In the model without additional complications, it is shown that the value of product-mix information increases with the number of items, whereas the value of global information decreases with the number of items. The value of both product-mix information and global information decreases with a budget constraint. Furthermore, the value of perfect information also decreases with a budget constraint. The probability of substitution decreases the value of product-mix information such that it is zero with complete substitution, and increases the value of global information so that it is equal to the value of perfect information with complete substitution.

1.4 Extensions

In this final section, we describe two recent extensions for handling the unconstrained MPNP for the case of substitute products. The first extension examines the case where the demand is price dependent. This is a common situation that arises in practices that customers substitute a different product when the price of the desired product has increased. For example, the supermarkets stock two different brand shampoos with same price and similar quality. If one of the products has increased their price, the price sensitive customers will choose the product with the lower price. The second case addresses the situation where demand is quantity dependent. This is reasonable in reality because an increase in shelf space for a product attracts more customers to buy it due to its visibility and popularity. Conversely, low stocks of certain goods (e.g., perishable food) might leave the impression that they are not fresh. In both cases, we present the results for a stylized scenario for two products.

1.4.1 Price Linear Demand

Carrillo et al. (2011) analyze the stocking decision under price linear demand for substitutive products. The demand function for product $i(i=1,2)$ is:

$$x_i = a_i - b_i p_i + s p_j + \varepsilon_i, \tag{1.44}$$

where a_i is the market share for product i, b_i represents the price elasticity of demand for product i, and s is the symmetric price-based substitution effect parameter. ε_i is defined as a continuous random variable with probability density function $f(\cdot)$ and cumulative distribution function $F(\cdot)$ in the range of $[-d_i, d_i]$ with mean μ_i The profit for each product is:

$$\pi_i(Q_i, x_i) = \begin{cases} p_i x_i - v_i Q_i + g_i(Q_i - x_i) & \text{if } Q_i \geq x_i \\ p_i Q_i - v_i Q_i - B_i(x_i - Q_i) & \text{if } Q_i < x_i \end{cases}. \tag{1.45}$$

Let $z_i = Q_i - a_i + b_i p_i$, the expected profit for each product i is:

$$\begin{aligned} E[\Pi(z_1, z_2, p_1, p_2)] = &\left\{ \int_{-d_1}^{z_1} [p_1(a_1 - b_1 p_1 + s p_2 + \varepsilon_1) + g_1(z_1 - \varepsilon_1)] f(\varepsilon_1) \mathrm{d}\varepsilon_1 \right\} \\ &+ \left\{ \int_{z_1}^{d_1} [p_1(a_1 - b_1 p_1 + s p_2 + z_1) + B_1(z_1 - \varepsilon_1)] f(\varepsilon_1) \mathrm{d}\varepsilon_1 \right\} \\ &- v_1(a_1 - b_1 p_1 + s p_2 + z_1) \\ &+ \left\{ \int_{-d_2}^{z_2} [p_2(a_2 - b_2 p_2 + s p_1 + \varepsilon_2) + g_2(z_2 - \varepsilon_2)] f(\varepsilon_2) \mathrm{d}\varepsilon_2 \right\} \\ &+ \left\{ \int_{z_2}^{d_2} [p_2(a_2 - b_2 p_2 + s p_1 + z_2) + B_2(z_2 - \varepsilon_2)] f(\varepsilon_2) \mathrm{d}\varepsilon_2 \right\} \\ &- v_2(a_2 - b_2 p_2 + s p_1 + z_2) \\ = &\sum_{i=1}^{2} \Bigg\{ (p_i - v_i)(a_i - b_i p_i) - (v_i - g_i) z_i \\ &- (p_i - g_i) \left[\int_{z_i}^{d_i} (\varepsilon_i - z_i) f(\varepsilon_i) \mathrm{d}\varepsilon_i - \mu_i \right] \Bigg\} + B_i \int_{z_i}^{d_i} (z_i - \varepsilon_i) \\ &\times f(\varepsilon_i) \mathrm{d}\varepsilon_i + s p_1(p_2 - v_2) + s p_2(p_1 - v_1). \end{aligned} \tag{1.46}$$

The FOCs are:

$$\frac{\partial E[\Pi]}{\partial z_1} = -v_1 + g_1 F(z_1) + (p_1 + B_1)[1 - F(z_1)], \tag{1.47}$$

$$\frac{\partial E[\Pi]}{\partial z_2} = -v_2 + g_2 + (p_2 + B_2)[1 - F(z_2)], \tag{1.48}$$

$$\frac{\partial E[\Pi]}{\partial p_1} = 2b_1\left[\frac{a_1 + b_1 v_1}{2b_1} - p_1\right] - \int_{z_1}^{d_1} (\varepsilon_1 - z_1) f(\varepsilon_1) \mathrm{d}\varepsilon_1 + 2sp_2 - sv_2 + \mu_1, \tag{1.49}$$

$$\frac{\partial E[\Pi]}{\partial p_2} = 2b_2\left[\frac{a_2 + b_2 v_2}{2b_2} - p_2\right] - \int_{z_2}^{d_2} (\varepsilon_2 - z_2) f(\varepsilon_2) \mathrm{d}\varepsilon_2 + 2sp_1 - sv_1 + \mu_2. \tag{1.50}$$

The second-order conditions are:

$$\frac{\partial^2 E[\Pi]}{\partial {z_i}^2} = (g_i - p_i - B_i) f(z_i) \qquad \text{for } i = 1, 2, \tag{1.51}$$

$$\frac{\partial^2 E[\Pi]}{\partial {p_i}^2} = -2b_i \qquad \text{for } i = 1, 2, \tag{1.52}$$

$$\frac{\partial^2 E[\Pi]}{\partial z_1 \partial p_1} = 1 - F(z_1), \tag{1.53}$$

$$\frac{\partial^2 E[\Pi]}{\partial z_1 \partial p_2} = 0, \tag{1.54}$$

$$\frac{\partial^2 E[\Pi]}{\partial z_2 \partial p_2} = 1 - F(z_2), \tag{1.55}$$

$$\frac{\partial^2 E[\Pi]}{\partial z_2 \partial p_1} = 0, \tag{1.56}$$

$$\frac{\partial^2 E[\Pi]}{\partial p_1 \partial p_2} = 2s. \tag{1.57}$$

We can't prove that the Hessian is strictly concave without the specific value for parameters.

For specific values of z_1 and z_2, (1.49) and (1.50) are strictly and jointly concave in p_1 and p_2. Since it was assumed that $b_j > s$ for $j = 1, 2.$, (1.52) and (1.57) indicate that $|H_1| < 0$ and $|H_2| = 4b_1 b_2 - 4s^2 > 0$. Thus, for given values of z_1 and z_2, the optimal prices can be determined by solving the following simultaneous equations (obtained by setting the FOCs in (1.49) and (1.50) equal to 0):

$$-2b_1 p_1 + 2sp_2 = \int_{z_1}^{d_1} (\varepsilon_1 - z_1) f(\varepsilon_1) \mathrm{d}\varepsilon_1 + sv_2 - (a_1 + b_1 v_1) - \mu_1, \tag{1.58}$$

$$2sp_1 - 2b_2 p_2 = \int_{z_2}^{d_2} (\varepsilon_2 - z_2) f(\varepsilon_2) \mathrm{d}\varepsilon_2 + sv_1 - (a_2 + b_2 v_2) - \mu_2. \tag{1.59}$$

The solution to this set of equations is:

$$p_1(z_1,z_2) = \frac{b_2u_1 + su_2}{2(b_1b_2 - s^2)}, \tag{1.60}$$

$$p_2(z_1,z_2) = \frac{b_1u_2 + su_1}{2(b_1b_2 - s^2)}, \tag{1.61}$$

where $u_1 = (a_1 + b_1c_1) - \int_{z_1}^{d_1}(\varepsilon_1 - z_1)f(\varepsilon_1)\mathrm{d}\varepsilon_1 - sc_2 + \mu_1$ and $u_2 = (a_2 + b_2c_2) - \int_{z_2}^{d_2}(\varepsilon_2 - z_2)f(\varepsilon_2)\mathrm{d}\varepsilon_2 - sc_1 + \mu_2$.

Then the following algorithm could determine the optimal prices, stocking quantities, and the corresponding optimal profit:

1. Set $z_1 = -d_1 - 0.01$; $z_2 = -d_2 - 0.01$; $Profit = 0$; $m_1 = m_2 = 0$, $p_{1t} = p_1 = 0$; $p_{2t} = p_2 = 0$ and $z_{1t} = z_{2t} = 0$.
2. $z_1 = z_1 + 0.01$. If $z_1 > d_1$, go to 6.
3. $z_2 = z_2 + 0.01$. If $z_2 > d_2$, go to 2.
4. Compute $p_1(z_1,z_2)$ using (1.60) and $p_2(z_1,z_2)$ using (1.61). Set $p_{1t} = p_1(z_1,z_2)$ and $p_{2t} = p_2(z_1,z_2)$.
5. Compute $E[\Pi(z_1,z_2,p_{1t},p_{2t}]$ using (1.46). If $Profit > E[\Pi(z_1,z_2,p_{1t},p_{2t})]$, go to 3, else set $Profit = E[\Pi(z_1,z_2,p_{1t},p_{2t})]$; $z_{1t} = z_1$; $z_{2t} = z_2$; $m_1 = p_{1t}$, $m_2 = p_{2t}$ and go to 3.
6. The optimal market prices are: $p_1* = m_1$ and $p_2* = m_2$; the optimal stocking quantities are: $q_1* = a_1 - b_1m_1 + sm_2 + z_{1t}$, and $q_2* = a_2 - b_2m_2 + sm_1 + z_{2t}$ and associated optimal profit is $Profit$.

1.4.2 Quantity Linear Demand

The demand function for quantity linear demand of product i ($i = 1,2$) is

$$x_i = a_i + b_iq_i - sq_j + \varepsilon_i, \tag{1.62}$$

where a_i represents the relative market share for product i, b_i is the quantity elasticity of demand, and s is the symmetric quantity-based substitution effect parameter. Also ε_i is defined as a continuous random variable with probability density function $f(\cdot)$ and cumulative distribution function $F(\cdot)$ in the range of $[-d_i,d_i]$ with mean μ_i. The profit for each product is:

$$E[\Pi_i(z_1,z_2,q_1,q_2)] = \left\{\int_{-d_i}^{z_i}[p_i(a_i + b_iq_i - sp_j + \varepsilon_i) + g_i(z_i - \varepsilon_i)]f(\varepsilon_i)\mathrm{d}\varepsilon_i\right\} + \left\{\int_{z_i}^{d_i}[p_i(a_i + b_iq_i - sq_j + z_i) + B_i(z_i - \varepsilon_i)]f(\varepsilon_i)\mathrm{d}\varepsilon_i\right\}$$

$$
\begin{aligned}
&-v_i(a_i + b_i q_i - s q_j + z_i) \\
&= (p_i - v_i)(a_i + b_i q_i - s q_j) - (v_i - g_i) z_i + (p_i - g_i)\mu_i \\
&-(p_i - g_i + B_i)\int_{z_i}^{d_i} (\varepsilon_i - z_i) f(\varepsilon_i)\,\mathrm{d}\varepsilon_i,
\end{aligned} \tag{1.63}
$$

where $z_i = q_i - (a_i + b_i q_i - s q_j)$. The total profit is $E[\Pi(z_1, z_2, q_1, q_2)] = E[\Pi_1)] + E[\Pi_2]$

The FOCs are:

$$
\begin{aligned}
\frac{\partial E[\Pi]}{\partial q_i} &= (p_i - g_i)b_i + (g_i - v_i) + (p_i - g_i + B_i)[1 - F(z_i)](1 - b_i) + \\
&-s\{(p_j - g_j) + (p_j - g_j + B_j)[1 - F(z_j)]\}.
\end{aligned} \tag{1.64}
$$

The second-order conditions are:

$$
\frac{\partial^2 E[\Pi]}{\partial {q_i}^2} = -(p_i - g_i + B_i)(1 - b_i)^2 f(z_i) - (p_j - g_j + B_j)s^2 f(z_j), \tag{1.65}
$$

$$
\frac{\partial^2 E[\Pi]}{\partial q_1 \partial q_2} = -(p_1 - g_1 + B_1)(1 - b_1)s f(z_1) - (p_2 - g_2 + B_2)(1 - b_2)s f(z_2). \tag{1.66}
$$

From the Hessian Matrix, $|H_1| < 0$, $|H_2| = u_1 u_2[(1 - b_1)(1 - b_2) - s^2]^2 > 0$, where $u_i = (p_i - g_i + B_i)f(y_i)$. Since the objective function is strictly concave, we can obtain the solution for this problem from the FOCs (i.e., by setting them equal to 0 and simultaneously solving for the decision variables).

1.5 Conclusions and Directions for Future Research

In this paper, we have reviewed and critiqued the literature to date for the MPNP. As is obvious, the majority of prior research has focused on determining the optimal stocking policy for the constrained MPNP. More recent work on exploring the impact of substitutability has also been undertaken and in this setting, we present two possible extensions of the MPNP for price and quantity substitution effects. For both these cases, we show that optimal solutions can be obtained through either a search process (for price substitution) or a structural analysis (for quantity substitution). Here we point out a few areas where further research is needed.

1.5.1 Price-Dependent Demand

Previous researches on MPNP assume the independence of price and market demand. Recent work has mitigated this issue by addressing the joint ordering and pricing problem in the MPNP framework. But most of these works only consider single budget constraint, while in practice the retailer may face multiple resource constraints. So it would be interesting to extend the study to consider the problem with multiple constraints. Due to the complexity of the problem, high quality heuristics procedures are anticipated to find good solutions.

1.5.2 Multiple Suppliers

Nearly all the models in this chapter assume single supplier. However, in practice, retailers may face several suppliers when making the merchandise decision. It would be interesting to incorporate multiple suppliers into MPNP, especially under price competition between potential suppliers and availability of several supply options. We see many opportunities for future research to help bridge this gap.

1.5.3 Product Substitution

Incorporating the substitution effects can have a significant effect on profitability. However, most previous studies on the substitution effects of the MPNP only focus on two products substitutability. It would be interesting to extend the analysis in a more generalized case. This extension requires a better understanding of interdependencies among the demands for related products. So an empirical investigation of the generalized substitution effects in customer decision making will also be an attractive future research area.

1.5.4 Risk and Hedging

The classical newsvendor problem is based on the assumption that most of the supply chains are risk neutral. The research on risk-averse supply chains has been considered by several authors. However, these papers focused on independent demand. As an extension, price and quantity-dependent demands can be considered. The hedging problem has been tackled by Vaagen and Wallace in the fashion industry, this research could be extended to other industries with different sales behavior. Also, the pricing strategies for these hedging portfolios could be examined to identify policies that further reduce profit risk.

References

Abdel-Malek, L., & Areeratchakul, N. (2007). A quadratic programming approach to the multi-product newsvendor problem with side constraints. *European Journal of Operational Research*, *176*(3), 1607–1619.

Abdel-Malek, L., & Montanari, R. (2005a). An analysis of the multi-product newsboy problem with a budget constraint. *International Journal of Production Economics*, *97*(3), 296–307.

Abdel-Malek, L., & Montanari, R. (2005b). On the multi-product newsboy problem with two constraints. *Computers & Operations Research*, 32, 2095–2116.

Abdel-Malek, L. Montanari, R., & Morales, L. (2004). Exact, approximate, and generic iterative models for the multi-product newsboy problem with budget constraint. *International Journal of Production Economics*, *91*(2), 189–198.

Anupindi, R., Dada, M., & Gupta, S. (1998). Estimation of consumer demand with stock-out based substitution: an application to vending machine products. *Marketing Science*, *17*(4), 406–423.

Areeratchakul, N., & Abdel-Malek, L. (2006). An approach for solving the multi-product newsboy problem. *International Journal of Operational Research*, *3*(3), 219–227.

Arrow, K., Harris, T., & Marshak, J. (1951). Optimal inventory policy. *Econometrica*, *19*(3), 250–272.

Aviv, Y., & Federgruen, A. (2001). Capacitated multi-item inventory systems with random and seasonally fluctuating demands: implications for postponement strategies. *Management Science*, *47*(4), 512–531.

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

Ben-Daya, M., & Raouf, A. (1993). On the constrained multi-tem single-period inventory problem. *International Journal of Operations and Production Management*, *13*(11), 104–112.

Cai, L., Chen, J., & Yan, H. (2004). Technical note: single-period two-product inventory model with substitution: solution and properties. *Journal of Systems Science and Systems Engineering*, *13*, 112–123.

Carrillo, J., Vakharia, A., & Wang, R. (2011). Environmental implications for online retailing. *Working Paper*.

Casimir, R. (2002). The value of information in the multi-item newsboy problem. *Omega*, *30*(1), 45–50.

Chen, L., & Chen, Y. (2010). A multiple-item budget-constraint newsboy problem with a reservation policy. *Omega*, *38*(6), 431–439.

Cheng, T.C.E., & Choi, T.M. (2010). Innovative quick response programs in logistics and supply chain management. *International handbooks on information systems*. Berlin: Springer.

Choi, S., Ruszczynski, A., & Zhao, Y. (2011). A multiproduct risk-averse newsvendor with law-invariant coherent measures of risk. *Operations Research*, *59*(2), 346–364.

Chung, C., Flynn, J., & Kirca, O. (2008). A multi-item newsvendor problem with preseason production and capacitated reactive production. *European Journal of Operational Research*, *188*(3), 775–792.

Dutta, P., & Chakraborty, D. (2010). Incorporating one-way substitution policy into the newsboy problem with imprecise customer demand. *European Journal of Operational Research*, *200*(1), 99–110.

Erlebacher, S. (2000). Optimal and heuristic solutions for the multi-item newsvendor problem with a single capacity constraint. *Production and Operations Management*, *9*(3), 303–318.

Gallego, G., & Moon, I. (1993). The distribution free newsboy problem: review and extensions. *The Journal of the Operational Research Society*, *44*(8), 825–834.

Hadley, G., & Whitin, T. (1963). *Analysis of inventory systems*. New jersey: Prentice-Hall.

Khouja, M., & Mehrez, A. (1996). A multi-product constrained newsboy problem with progressive multiple discounts. *Computers and Industrial Engineering*, *30*(1), 95–101.

Khouja, M., Mehrez, A., & Rabinowitz, G. (1996). A two-item newsboy problem with substitutability. *International Journal of Production Economics*, *44*(3), 267–275.

Lau, A., & Lau, H. (1988). Maximizing the probability of achieving a target profit in a two-product newsboy problem. *Decision Sciences*, *19*(2), 392–409.

Lau, H., & Lau, A. (1995). The multi-product multi-constraint newsboy problem: applications, formulation and solution. *Journal of Operations Management*, *13*(2), 153–162.

Lau, H., & Lau, A. (1996). The newsstand problem: a capacitated multiple-product single-period inventory problem. *European Journal of Operational Research*, *94*(1), 29–42.

Lau, H., & Lau, A. (1997). Some results on implementing a multi-item multi-constraint single-period inventory model. *International Journal of Production Economics*, *48*(2), 121–128.

Mahajan, S., & van Ryzin, G. (2001). Stocking retail assortments under dynamic consumer substitution. *Operations Research*, *49*(3), 334–351.

McGillivray, A. R., & Silver, E. A. (1978). Some concepts for inventory control under substitutable demand. *INFOR*, *16*(1), 47–63.

Mieghem, J. V., & Rudi, N. (2002). Newsvendor networks: inventory management and capacity investment with discretionary activities. *Manuacturing Service Operations Management*, *4*(4), 313–335.

Moon, I., & Silver, E. A. (2000). The multi-item newsvendor problem with a budget constraint and fixed ordering costs. *The Journal of the Operational Research Society*, *51*(5), 602–608.

Nagarajan, M., & Rajagopalan, S. (2008). Inventory models for substitutable products: optimal policies and heuristics. *Management Science*, *54*(8), 1453–1466.

Nahmias, S., & Schmidt, C. P. (1984). An efficient heuristic for the multi-item newsboy problem with a single constraint. *Naval Research Logistics Quarterly*, *31*(3), 463–474.

Netessine, S., & Rudi, N. (2003). Centralized and competitive inventory models with demand substitution. *Operations Research*, *51*(2), 329–335.

Niederhoff, J. (2007). Using separable programming to solve the multi-product multiple ex-ante constraint newsvendor problem and extensions. *European Journal of Operational Research*, *176*(2), 941–955.

Ozler, A., Tan, B., & Karaesmen, F. (2009). Multi-product newsvendor problem with value-at-risk considerations. *International Journal of Production Economics*, *117*(2), 244–255.

Parlar, M., & Goyal, S. (1984). Optimal ordering decisions for two substitutable products with stochastic demands. *OpSearch*, 21, 1–15.

Pasternack, B. A., & Drezner, Z. (1991). Optimal inventory policies for substitutable commodities with stochastic demand. *Naval Research Logistics (NRL)*, *38*(2), 221–240.

Qin, Y., Wang, R., Vakharia, A., Chen, Y., & Hanna-Seref, M. (2011). The newsvendor problem: review and directions for future research. *European Journal of Operational Research*, *213*, 361–374.

Rajaram, K., & Tang, C. (2001). The impact of product substitution on retail merchandising. *European Journal of Operational Research*, *135*(3), 582–601.

Shah, J., & Avittathur, B. (2007). The retailer multi-item inventory problem with demand cannibalization and substitution. *International Journal of Production Economics*, *106*(1), 104–114.

Shao, Z., & Ji, X. (2006). Fuzzy multi-product constraint newsboy problem. *Applied Mathematics and Computation*, *180*(1), 7–15.

Shi, J., & Zhang, G. (2010). Multi-product budget-constrained acquisition and pricing with uncertain demand and supplier quantity discounts. *International Journal of Production Economics*, *128*(1), 322–331.

Shi, J., Zhang, G., & Sha, J. (2011). Jointly pricing and ordering for a multi-product multi-constraint newsvendor problem with supplier quantity discounts. *Applied Mathematical Modelling*, *35*(6), 3001–3011.

Silver, E., Pyke, D., & Peterson, R. (1998). *Inventory management and production planning and scheduling*. New York: Wiley.

Smith, S., & Agrawal, N. (2000). Management of multi-item retail inventory systems with demand substitution. *Operations Research*, *48*(1), 50–64.

Vaagen, H., & Wallace, S. W. (2008). Product variety arising from hedging in the fashion supply chains. *International Journal of Production Economics*, *114*(2), 431–455.

Vairaktarakis, G. (2000). Robust multi-item newsboy models with a budget constraint. *International Journal of Production Economics*, *66*(3), 213–226.

Zhang, B., & Du, S. (2010). Multi-product newsboy problem with limited capacity and outsourcing. *European Journal of Operational Research*, *202*(1), 107–113.

Zhang, B., & Hua, Z. (2010). A portfolio approach to multi-product newsboy problem with budget constraint. *Computers & Industrial Engineering*, *58*, 759–765.

Zhang, B., Xu, X., & Hua, Z. (2009). A binary solution method for the multi-product newsboy problem with budget constraint. *International Journal of Production Economics*, *117*(1), 136–141.

Zhang, G. (2010). The multi-product newsboy problem with supplier quantity discounts and a budget constraint. *European Journal of Operational Research*, *206*(2), 350–360.

Chapter 2
A Multi-item Risk-Averse Newsvendor with Law Invariant Coherent Measures of Risk

Sungyong Choi

Abstract I consider a multi-product risk-averse newsvendor under the law-invariant coherent measures of risk. I first establish a few fundamental properties of the model regarding the convexity of the problem and the symmetry of the solution, and study the impacts of risk aversion and shift in mean demand to the optimal solution with independent demands. Specifically, I show that for identical products with independent demands, increased risk aversion leads to decreased orders. For a large but finite number of heterogenous products with independent demands, I derive closed-form approximations for the optimal order quantities. The approximations are as simple to compute as the classical risk-neutral solutions. I also show that the risk-neutral solution is asymptotically optimal as the number of products tends to be infinity, and thus risk aversion has no impact in the limit. For a risk-averse newsvendor with dependent demands, I show that positively (negatively) dependent demands lead to a lower (higher) optimal order quantities than independent demands. Using a numerical study, I examine the convergence rates of the approximations and develop additional insights on the interplay between dependent demands and risk aversion.

Keywords Multiple products • Newsvendor • Risk-averse • Coherent risk measures • Diversification • Portfolio

---

S. Choi (✉)
Department of Business Administration, College of Government and Business, Yonsei University, Wonju, Gangwon-Do, 220–710, Republic of Korea
e-mail: sungyongchoi@gmail.com; sungyong.choi@yonsei.ac.kr

T.-M. Choi (ed.), *Handbook of Newsvendor Problems*, International Series in Operations Research & Management Science 176, DOI 10.1007/978-1-4614-3600-3_2,

2.1 Introduction

2.1.1 Motivation

The multi-product newsvendor model is a classical model in the inventory control literature. In this model, there are multiple products to be sold in a single selling season. On the one hand, when demand exceeds supply for any product, the excessive demand is lost. On the other hand, when supply exceeds demand, the excessive inventory is sold at a loss. The firm's objective is to determine the optimal order quantity for each product so as to maximize a certain performance measure. This model finds its applications in many manufacturing, distribution, and retailing firms that handle short life cycle products.

The literature of the multi-product newsvendor model has mainly used risk-neutral performance measures as an objective function. For example, the company optimizes the expected average profit or average cost per product. Under these objective functions, the model is decomposable and one can consider each product separately as multiple single-product newsvendor models, unless resource constraints are imposed nor demand substitution is allowed. Under risk-averse objective functions, however, the model is generally not decomposable. One needs to consider all products simultaneously, as a portfolio.

Below, I first review the literature of risk-neutral multi-product inventory models by ways products interact. Then, I review the literature of risk models and its recent applications in supply chain inventory management.

Hadley and Whitin (1963) consider a multi-product newsvendor model with storage capacity or budget constraints, and provide the solution methods based on Lagrangian multiplier. Porteus (1990) presents a thorough review of various newsvendor models. Veinott (1965) considers the dynamic version of the multi-product inventory models in a multi-period setting, with general assumptions in demand process, cost parameters, and lead times. Conditions under which myopic policy is optimal are identified. Ignall and Veinott (1969) and Heyman and Sobel (1984) extend the work by identifying new conditions for the myopic policy in models with risk-neutral assumption, see Aviv and Federgruen (2001), Decroix and Arreola-Risa (1998), Evans (1967), and Federgruen (1984) for exact analysis and approximations. Other than resource constraints, multi-product newsvendor models are also studied under demand substitution, where unsatisfied demand of one product can be satisfied by on-hand inventory of another product. I refer to van Ryzin and Mahajan (1999) for a review on multi-item inventory systems with substitution.

My aim is to replace the risk-neutral performance measure by measures taking risk aversion into account. Such a model is generally not decomposable, and one needs to consider all products simultaneously, as a portfolio. In this paper, I lay the foundations of the multi-product newsvendor model under coherent measures of risk and derive its basic properties. They provide insight into the impact of risk aversion on the multi-product newsvendor with either independent or dependent demands. Moreover, I study asymptotic properties of the solution as the number of products tends to infinity and develop simple yet accurate approximations of risk-averse solutions, which allow fast computation of large-scale problems.

Below, I first review the literature of risk measures and their recent applications in supply chain inventory management. Then, I summarize my model and main results.

2.1.2 Risk Measures

The risk-neutral inventory models provide the best decision *on average*. This may be justified by the Law of Large Numbers. However, one cannot always rely on repeated similar chances. The first few outcomes may turn out to be very bad and entail unacceptable losses. Schweitzer and Cachon (2000) provide experimental evidence suggesting that inventory managers may be risk-averse for high-value products. Because of these reasons, attempts to overcome the drawbacks of the expected value optimization have a long history and there exist four typical approaches to model decision making under risk. They are expected utility theory, stochastic dominance, chance constraints, and mean-risk analysis. These approaches are related and consistent to some extent.

The expected utility theory of von Neumann and Morgenstern (1944) derives, from simple axioms, the existence of a nondecreasing *utility function*, which transforms in a nonlinear way the observed outcomes. The decision maker optimizes, instead of the expected outcome, the expected value of the utility function. In the maximization context, when the outcome represents profit, risk-averse decision makers have concave and nondecreasing utility functions.

The second approach is based on the theory of *stochastic dominance*, developed in statistics and economics (see Lehmann 1955; Hadar and Russell 1969 and references therein). Stochastic dominance relations are partial orders on the space of distributions, and thus allow for pairwise comparison of different solutions. An important feature of the stochastic dominance theory is its universal character with respect to utility functions. More specifically, the distribution of a random outcome V is preferred to random outcome Y in terms of a stochastic dominance relation if and only if expected utility of V is preferred to expected utility of Y for all utility functions in a certain class, called the generator of the relation. In particular, the second-order stochastic dominance corresponds to all concave nondecreasing utility functions, and is thus well suited to model risk-averse preferences. For an overview of these issues, see Müller and Stoyan (2002) and Levy (2006). Unfortunately, the stochastic dominance approach does not provide a simple computational recipe. In fact, it is a multiple criteria model with a continuum of criteria. Therefore, it has been used as a constraint (see Dentcheva and Ruszczyński 2003), and also utilized as a reference standard whether a particular solution approach is appropriate (see Ogryczak and Ruszczyński 1999; Ruszczyński and Vanderbei 2003).

Table 2.1 A counterexample to show problems of mean-variance models

	Policy 1	Policy 2
"Bad" Stake (0.5)	-1	-1
"Good" Stake (0.5)	1	3
Mean (μ)	0	1
Variance (σ^2)	1	4
Absolute semi-deviation (σ_1)	1/2	1
Standard semi-deviation (σ_2)	$1/\sqrt{2}$	$\sqrt{2}$
$-\mu + \mathbb{1} \cdot \sigma^2$	1	3
$-\mu + \mathbb{1} \cdot \sigma_1$	1/2	0
$-\mu + \mathbb{1} \cdot \sigma_2$	$1/\sqrt{2}$	$\sqrt{2} - 1$

The third approach specifies constraints on probabilities of unfavorable events. Prékopa (2003) provides a thorough overview of the state of the art of the optimization theory with chance constraints. Theoretically, a chance constraint is a relaxed version of the stochastic dominance relation of the first-order, and thus it is related to the expected utility theory, but there is no equivalence. In finance, chance constraints are known under the name of Value-at-Risk (VaR) constraints. Chance constraints sometimes lead to nonconvex formulations of the resulting optimization problems.

The fourth approach, originating from finance, is the *mean-risk analysis*. It quantifies the problem in a lucid form of two criteria: the *mean* (the expected value of the outcome), and the *risk* (a scalar measure of the variability of the outcome). In the maximization context, one selects from the universe of all possible solutions those that are *efficient*: for a given value of the mean they minimize the risk, or equivalently, for a given value of risk they maximize the mean. Such an approach has many advantages: it allows one to formulate the problem as a parametric optimization problem, and it facilitates the trade-off analysis between mean and risk.

In the context of portfolio optimization, Markowitz (1959) used the variance of the return as the risk. It is easy to compute, and it reduces the financial portfolio selection problem to a parametric quadratic programming problem. One can, however, construct simple counterexamples that show the imperfection of the variance as the risk measure: it treats over-performance equally as under-performance, and more importantly it may suggest a portfolio which is stochastically dominated by another portfolio. Table 2.1 below summarizes a defect of mean-variance models. In Table 2.1, let me consider two policies, policy 1 and 2, defined at the two equally likely events, "Bad" and "Good." Then, policy 2 is stochastically bigger than policy 1. Here, both $-\mu + \sigma_1$ and $-\mu + \sigma_2$ are coherent risk measures. Then, with these two risk measures, policy 2 is preferred to policy 1, which shows consistency with stochastic dominance relations. However, with a mean-variance model, policy 1 may be preferred to policy 2 implying contradiction to stochastic dominance.

To overcome the drawbacks of the mean-variance analysis, the general theory of *coherent measures of risk* was suggested by Artzner et al. (1999) and extended to general probability spaces by Delbaen (2002). For further generalizations, see Föllmer and Schied (2002, 2004), Kusuoka (2003), Ruszczyński and Shapiro

(2005) and Ruszczyński and Shapiro (2006a). Dynamic version for a multi-period case were analyzed, among others, by Riedel (2004), Kusuoka and Morimoto (2004), Cheridito et al. (2006) and Ruszczyński and Shapiro (2006b). In this theory, an integrated performance measure is proposed, comprising both the mean and variability measures, and four axioms (Convexity, Monotonicity, Translation Equivariance, and Positive Homogeneity; see Sect. 2.3 for a precise definition) are imposed. Coherent measures of risk are extensions of the mean-risk analysis. It is known that coherent measures of risk are consistent with the 1st and 2nd order stochastic dominance relations (see Shapiro et al. 2009).

More specifically, in a multi-product newsvendor problem, these four axioms have following implications to guarantee consistency with intuition about rational risk-averse decision making. Thus, by satisfying the axioms, a coherent risk measure has certain attractive features, as compared to these measures, making it worth considering. First, Convexity axiom means that the global risk of a portfolio should be equal or less than the sum of its partial risks. Thus, this axiom is consistent with diversification effects. Second, Monotonicity axiom is consistent with the first-order stochastic dominance relation. Third, Translation Equivariance axiom means that adding a constant cost is equivalent to increasing the vendors performance measure by the same amount. On the contrary, adding a constant gain is equivalent to decreasing the vendors performance measure by the same amount. Therefore, by excluding the impact of constant gains or losses, fixed parts can be separated equivalently from the vendors random performance measure at every possible state of nature. Lastly, Positive Homogeneity axiom guarantees that the optimal solution does not change to rescaling of units.

Among the four axioms aforementioned, expected utility models and coherent risk measures share the properties of convexity and consistency with stochastic dominance. In addition, the coherent risk measures satisfy the axioms of Translation Equivariance and Positive Homogeneity. However, under expected utility theory, these two axioms typically do not hold; see, e.g., the exponential utility function in Howard (1988).

For inventory systems where the initial endowment effect is significant, i.e., when the initial wealth could affect the decision of a risk-averse manager, or when constant demand for some products could affect order quantities of other products, an expected utility model may be preferred to a model with a coherent risk measure, because the latter ignores the endowment effect. In newsvendor models, where inventory managers are mainly concerned about the overage and underage costs associated with random demand, and in other problems, where risk is primarily associated with uncertainty, coherent risk measures may capture risk preferences better. The following arguments speak in favor of coherent measures of risk: (1) Translation Equivariance allows them to properly rank risky alternatives by excluding the impact of constant gains or losses (see Artzner et al. 1999). (2) The Positive Homogeneity axiom ensures that their attitude to risk will not change when the unit system is changed (e.g., from dollars to cents). More importantly, this axiom indicates no diversification effect when demands are completely correlated. To see this, it is well known that the subadditivity property, $\rho(X+Y) \leq \rho(X) + \rho(Y)$, implies $\rho(nX) \leq n\rho(X)$. However $\rho(nX) < n\rho(X)$ would imply diversification effect even when the random demands are completely correlated. To avoid this counter-intuitive effect, it is left with $\rho(nX) = n\rho(X)$ which is the Positive Homogeneity axiom.

Several modifications and extensions of coherent measures of risk have been suggested in the literature, including *convex measures of risk, insurance risk measures, natural risk statistic*, and *tradeable measures of risk*. I point out that all these risk measures ignore the initial endowment effect, implying consistency with Translation Equivariance.

Föllmer and Schied (2002) consider convex measures of risk, in which the Positive Homogeneity axiom is relaxed. Again, in my context, this may lead to a diversification effect when demands are completely correlated; it may also lead to counterintuitive effects of changing risk attitudes when the outcomes are rescaled, by changing the currency in which profits are calculated, or by considering the average profit per product.

The other three risk measures do not satisfy the convexity axiom in general. They are based on the reality of financial markets where noncoherent risk measures, such as VaR (Value-at-Risk), are widely used. Wang et al. (1997) suggest insurance risk measures which are law invariant, and satisfy the axioms of conditional state independence, monotonicity, comonotone additivity and continuity. Heyde (2006) propose the natural risk statistics, which is also law invariant, and in which the convexity axiom is required only for comonotone random variables. Ahmed et al. (2008) show that such a risk measure can be represented as a composition of a coherent measure of risk and a certain law preserving transformation, and thus the insights into models with coherent measures of risk are relevant for natural risk statistics. Pospišil et al. (2008) propose *tradeable measures of risk*. They argue that the proper risk measures should be constructed by historically realized returns. When compared to the coherent measures of risk, these risk measures appear to be much more difficult to handle, due to nonconvexity and/or nondifferentiability of the resulting model. I shall see that even in the case of coherent measures of risk the technical difficulties are substantial.

2.1.3 Risk-Averse Inventory Models

In recent years, risk-averse inventory models have received increasing attention in the supply chain management literature. Table 2.2 classifies the literature by inventory models and risk measures. Because there is no research so far directly applying stochastic dominance to this field, I drop it from the table.

Most work to date dealt with single-product inventory models. For newsvendor models, research focused on finding the optimal solution under a risk-averse measure, and studying the impact of the degree of risk aversion (among other model parameters) on the optimal solution. A typical finding is that as the degree of risk aversion increases, the optimal order quantity tends to decrease.

Table 2.2 Summary of the literature on risk-averse inventory models

	Single-product, single-period	Single-product, multi-period	Multi-product, single-period	Multi-echelon or multi-agent
Utility function	Lau (1980), Eeckhoudt et al. (1995), Agrawal and Seshadri (2000a), Gaur and Seshadri (2005)	Bouakiz and Sobel (1992), Chen et al. (2007)	van Mieghem (2007)	Agrawal and Seshadri (2000b), van Mieghem (2003), Gan et al. (2004)
Coherent measures of risk	Gotoh and Takano (2007), Ahmed et al. (2007), Choi and Ruszczyński (2008), Chen et al. (2009)	Ahmed et al. (2007)	Ağrali and Soylu (2006), Tomlin and Wang (2005), this paper	None
Mean-variance or mean-standard deviation	Anvari (1987), Chung (1990), Chen and Federgruen (2000), Gaur and Seshadri (2005), Martinez-de-Albeniz and Simchi-Levi (2006)	None	van Mieghem (2007)	Lau and Lau (1999), Tsay (2002), van Mieghem (2003), Gan et al. (2004)
Chance constraints or VaR	Lau (1980)	None	None	Gan et al. (2005)

For single-product but multi-period dynamic inventory models under risk aversion, the literature focuses on characterizing the structure of the optimal ordering or pricing policies and quantifying the impact of the degree of risk aversion on the optimal polices. Chen et al. (2006) review results in this direction.

For multi-product risk-averse newsvendor models, Tomlin and Wang (2005) study how characteristics of products (e.g., profit margin, demand correlation), resource reliability and firm's risk attitude affect the preference of resource flexibility and supply diversification. Under a downside risk measure and Conditional Value at Risk (CVaR), they show that for a risk-averse firm with unreliable resources, a supply chain can prefer dedicated resources than a flexible resource even if the cost of the latter is smaller than the former.

Newsvendor networks are studied by van Mieghem (2007), with many products and many resources under mean-variance and utility function approaches. The networks feature resource diversification, flexibility (e.g., ex post inventory capacity allocation) and/or demand pooling. The paper addresses the question of how the aforementioned operational strategies reduce total risk and create value. It shows that a risk-averse newsvendor may invest more resources in certain networks than a risk-neutral newsvendor (i.e., operational hedging) because such resources may reduce the profit variance and mitigate risk in the network. Among the three networks, the dedicated one is mostly related to my model. In this network, there are two products with correlated demand. The author characterizes the impact of demand correlation on the optimal order quantities in two extreme cases of complete positive or negative correlation. A numerical study is conducted to cover cases other than the extreme ones.

Finally, Ağrali and Soylu (2006) conduct a numerical investigation on a two-product newsvendor model under the risk measure of CVaR . Assuming a discretized multi-variate normal demand distribution, the authors studied the sensitivity of the optimal solution with respect to the mean and variance of demand, demand correlation, and various cost parameters. Interestingly, the report shows that as the demand correlation increases, the optimal order quantities tend to decrease.

For multi-echelon or multi-agent models, so far all papers consider single-product and single-period models. Lau and Lau (1999) study a manufacturer's pricing strategy and return policy under the mean-variance risk measure. Agrawal and Seshadri (2000b) introduce a risk-neutral intermediaries to offer mutually beneficial contracts to risk-averse retailers. Tsay (2002) studies how a manufacturer can use return policies to share risk under the mean-standard deviation measure. Gan et al. (2004) study Pareto-optimality for suppliers and retailers under various risk-averse measures. Gan et al. (2005) design coordination schemes of buyback and risk-sharing contracts in a supply chain under a Value-at-Risk constraint. For a review of the literature on risk aversion in capacity investment models and on operational hedging, see van Mieghem (2003).

2.1.4 My Model and Main Results

This paper considers a multi-product risk-averse newsvendor using a law-invariant coherent risk measure (see Sects. 2.2 and 2.3). As I argued in Sect. 2.1.2, coherent risk measures can be more attractive than the expected utility theory in the multi-product newsvendor problem due to their properties of Translation Equivariance and Positive Homogeneity.

The model presents a considerable challenge, both analytically and computationally, because the objective function cannot be decomposed by each product and one has to look at the totality of all products as a portfolio. In particular, one has to characterize the impact of risk aversion and demand dependence on the optimal solution, identify efficient ways to find the optimal solution, and connect this model to the financial portfolio theory. While Tomlin and Wang (2005) study a two-product system under CVaR, their focus is on the design of material flow topology and thus is different from mine.

I should also point out that in most practical cases where this model is relevant (either manufacturing or retailing), firms may have a large number of heterogenous products. Due to the complex nature of risk optimization models, they become practically intractable for problems of these dimensions. Thus, it is theoretically interesting and practically useful to study the asymptotic behavior of the system as the number of products tends to infinity and obtain fast approximation for large-size problems.

This work contributes to literature in the following ways: I first establish a few fundamental properties regarding the convexity of the model and the symmetry of the solution for the model in Sect. 2.4, and study the impacts of risk aversion and shift in mean demand to the optimal solution with independent demands in Sect. 2.5. I then consider large but finite number of independent heterogenous products, for which I develop closed-form approximations in Sect. 2.6 which are exact in the single-product case. The approximations are as simple to compute as the risk-neutral solutions. I also show that under certain regularity conditions, the risk-neutral solutions are asymptotically optimal under risk aversion, as the number of products tends to be infinity. This asymptotic result has an important economic implication: companies with many products or product families with low demand dependence need to look only at risk-neutral solutions, even if they are risk-averse.

The impact of dependent demands under risk aversion poses a substantial analytical challenge. By utilizing the concept of associated random variables, I prove in Sect. 2.7 that in a risk-averse two-product model with positively dependent demands the optimal order quantities are lower than for independent demands, while for negatively dependent demands the optimal order quantities are higher. Using a sample-based optimization, I conduct in Sect. 2.8 a numerical study, which demonstrates that the approximations converge quickly to the optimal solutions as the number of products increases. It also provides additional insights into the impact of dependent demands. Specifically, I identify counterexamples to show that increased risk aversion can lead to greater optimal order quantities for strongly negatively dependent demands. In Sect. 2.9, I summarize the paper and compare the multi-product risk-averse newsvendor model to the financial portfolio problem.

2.2 Problem Formulation

Given products $j = 1,\dots,n$, let $x = (x_1, x_2, \dots, x_n)$ be the vector of ordering quantities and let $D = (D_1,\dots,D_n)$ be the demand vector. I also define $r = (r_1,\dots,r_n)$ to be the price vector, $c = (c_1,\dots,c_n)$ to be cost vector, and $s = (s_1,\dots,s_n)$ to be the vector of salvage values. Finally, let $f_j(\cdot)$ and $F_j(\cdot)$ be the marginal probability density function (pdf), if it exists, and the marginal cumulative distribution function (cdf) of D_j, respectively. Denote $\bar{F}_j(\xi) = 1 - F_j(\xi)$.

Setting $\bar{c}_j = c_j - s_j$ and $\bar{r}_j = r_j - s_j$, I can write the profit function as follows:

$$\Pi(x,D) = \sum_{j=1}^{n} \Pi_j(x_j, D_j), \tag{2.1}$$

where

$$\begin{aligned} \Pi_j(x_j, D_j) &= -\bar{c}_j x_j + \bar{r}_j \min\{x_j, D_j\} \\ &= (r_j - c_j)x_j - (r_j - s_j)(x_j - D_j)^+, \quad j = 1,\dots,n, \end{aligned} \tag{2.2}$$

with $(x)^+ = \max\{x, 0\}$. I assume that the demand vector D is random and nonnegative. Thus, for every $x \geq 0$ the profit $\Pi(x,D)$ is a real bounded random variable.

The risk-neutral multi-product newsvendor optimization problem is to maximize the expected profit:

$$\max_{x\geq 0} \mathbb{E}[\Pi(x,D)]. \tag{2.3}$$

This problem can be decomposed into independent problems, one for each product. Thus, under risk neutrality, a multi-product newsvendor problem is equivalent to multiple single-product newsvendor problems. However, as I have mentioned it in the introduction, this formulation is inappropriate, if one is concerned with few (or just one) realizations and the Law of Large Numbers cannot be invoked.

Under a coherent risk measure, the optimization problem of the risk-averse newsvendor is defined as follows:

$$\min_{x\geq 0} \rho[\Pi(x,D)], \tag{2.4}$$

where $\rho[\cdot]$ is a law-invariant coherent measure of risk, and $\Pi(x,D)$ represents the profit of the newsvendor, as defined in (2.1). It is worth stressing that problem (2.4) cannot be decomposed into independent subproblems, one for each product. Thus, it is necessary to consider the portfolio of products as a whole.

2.3 Coherent Measures of Risk

I present a formal definition of the coherent measures of risk following the abstract approach of Ruszczyński and Shapiro (2005, 2006a). Let $(\Omega, \mathscr{F})$ be a certain measurable space. In my case, Ω is the probability space on which D is defined. An uncertain outcome (in my case, $\Pi(x,D)$) is represented by a measurable function $V : \Omega \to \mathbb{R}$. I specify the vector space $\mathscr{X}$ of possible functions; in my case it is sufficient to consider $\mathscr{X} = \mathscr{L}_\infty(\Omega, \mathscr{F}, P)$, which is the space of all bounded measurable functions on $[0,1]$. Indeed, for a fixed order quantity x, the function $\omega \to \Pi(x, D(\omega))$ is bounded. For any V and $Y \in \mathscr{X}$, I write $V \succeq Y$ if $V \geq Y$ almost surely (or with probability 1).

In the minimization context, a *coherent measure of risk* is a function $\rho : \mathscr{X} \to \mathbb{R}$ satisfying the following axioms:

Convexity: $\rho(\alpha V + (1-\alpha)Y) \leq \alpha\rho(V) + (1-\alpha)\rho(Y)$, for all $V, Y \in \mathscr{X}$ and all $\alpha \in [0,1]$.

Monotonicity: If $V, Y \in \mathscr{X}$ and $V \succeq Y$, then $\rho(V) \leq \rho(Y)$.

Translation Equivariance: If $a \in \mathbb{R}$ and $V \in \mathscr{X}$, then $\rho(V+a) = \rho(V) - a$.

Positive Homogeneity: If $t \geq 0$ and $V \in \mathscr{X}$, then $\rho(tV) = t\rho(V)$.

A coherent measure of risk $\rho(\cdot)$ is called *law invariant*, if the value of $\rho(V)$ depends only on the distribution of V, that is, $\rho(V_1) = \rho(V_2)$ if V_1 and V_2 have identical distributions. It implies that only the distribution matters, but not particular realizations. This axiom may look so natural. However, each random variable is actually defined by probability distribution as well as the field of events with a sigma-algebra structure. Although all practical risk measures are all law invariant, it is theoretically possible to construct a non law-invariant risk measure. From now on, without loss of generality, "coherent risk measures" actually mean "law-invariant coherent risk measures" unless mentioned explicitly. For more details of mathematical properties of law invariance, see Acerbi and Tasche (2002), Delbaen (2002) and Kusuoka (2003).

Important examples of law-invariant coherent measures of risk are obtained from mean–risk models of form:

$$\rho(V) = -\mathbb{E}[V] + \lambda r[V], \tag{2.5}$$

where $\lambda > 0$ and $r[\cdot]$ is a variability measure of the random outcome V. Popular examples of $r[\cdot]$ are the *semideviation* of order $p \geq 1$:

$$\sigma_p[V] = \mathbb{E}\left[\{(\mathbb{E}[V] - V)^+\}^p\right]^{\frac{1}{p}}, \tag{2.6}$$

and *weighted mean-deviation from quantile*:

$$r_\beta[V] = \min_{\eta \in \mathbb{R}} \mathbb{E}\left[\max\left((1-\beta)(\eta - V), \beta(V - \eta)\right)\right], \quad \beta \in (0,1). \tag{2.7}$$

The optimal η in the problem above is the β-quantile of V. Optimization models with (2.6) and (2.7) were considered in Ogryczak and Ruszczyński (1999, 2001, 2002). In the maximization context, from the practical point of view, it is most reasonable to consider $\beta \in (0, 1/2]$, because then $r_\beta[V]$ penalizes the left tail of the distribution of V much higher than the right tail.

The equation $\rho[\cdot]$ defined at (2.5), with $r[\cdot] = \sigma_p[\cdot]$ and $p \geq 1$, is a coherent measure of risk, provided that $\lambda \in [0, 1]$. When $r[\cdot] = r_\beta[\cdot]$, (2.5) is a coherent measure of risk, if $\lambda \in [0, 1/\beta]$. All these results can be found in Ruszczyński and Shapiro (2006a).

The mean-deviation from quantile $r_\beta[\cdot]$ is connected to the *Average Value-at-Risk* (AVaR), also known as *expected shortfall* or CVaR in Rockafellar and Uryasev (2000), as follows:

$$\text{AVaR}_\beta(V) = -\max_{\eta \in \mathbb{R}} \left\{ \eta - \frac{1}{\beta} \mathbb{E}\left[(\eta - V)^+ \right] \right\} = -\mathbb{E}[V] + \frac{1}{\beta} r_\beta[V]. \tag{2.8}$$

All these relations can be found in Föllmer and Schied (2004), Ogryczak and Ruszczyński (2002) and Ruszczyński and Vanderbei (2003) (with obvious adjustments for the sign of V). The relation (2.8) allows me to interpret $\text{AVaR}_\beta(V)$ as a special case of the mean–risk model where $r[V]$ is a deviation from quantile in (2.7) with $\lambda = 1/\beta$.

One of the fundamental results in the theory of law-invariant measures is the theorem of Kusuoka (2003): *For every lower semicontinuous law-invariant coherent measure of risk* $\rho[\cdot]$ *on* $\mathscr{L}_\infty(\Omega, \mathscr{F}, P)$*, with an atomless probability space* $(\Omega, \mathscr{F}, P)$*, there exists a convex set* $\mathscr{M}$ *of probability measures on* $(0, 1]$ *such that*

$$\rho[V] = \sup_{\mu \in \mathscr{M}} \int_0^1 \text{AVaR}_\beta[V]\, \mu(\mathrm{d}\beta). \tag{2.9}$$

Using identity (2.8), I can rewrite $\rho[V]$ as follows:

$$\rho[V] = -\mathbb{E}[V] + \sup_{\mu \in \mathscr{M}} \int_0^1 \frac{1}{\beta} r_\beta[V]\, \mu(\mathrm{d}\beta). \tag{2.10}$$

This means that every problem (2.4) with a coherent law-invariant measure of risk is a mean–risk model, with the variability measure

$$\varkappa_{\mathscr{M}}[V] = \sup_{\mu \in \mathscr{M}} \int_0^1 \frac{1}{\beta} r_\beta[V]\, \mu(\mathrm{d}\beta). \tag{2.11}$$

To illustrate the impact of scaling (the unit system) on risk measurement, I compare solutions of a single-product risk-averse newsvendor model under the

Table 2.3 The impact of rescaling on solutions—a coherent measure of risk, entropic exponential utility function, and a mean–variance model

	Unit of profit measurement				
	1 Dollar	30 Cents	10 Cents	3 Cents	1 Cent
Coherent	20.7824	20.7824	20.7824	20.7824	20.7824
Entropic exponential	20.7786	17.0952	12.2944	7.2879	4.8568
Mean-variance	20.7918	17.6962	14.4454	11.4197	9.5603

coherent risk measure with (2.5) and (2.7), the entropic exponential utility function $\frac{1}{\lambda}\ln\mathbb{E}\left[\mathrm{e}^{-\lambda\Pi(x;D)}\right]$ and the mean-variance model. The entropic exponential utility function is an example of a convex measure of risk which is not coherent and is equivalent to an exponential utility function by a certainty equivalent operator.

I select parameters for each risk measure so that they have the same optimal solution when the unit of profit measurement is one dollar. Specifically, I set $r = 15$, $c = 10$, and $s = 7$ (in dollars) for all three risk measures. Demand follows a lognormal distribution with $\mu = 3$ and $\sigma = 0.4724$. This demand distribution is used in all instances. For the coherent measure of risk, I set $\beta = 0.5$ and $\lambda = \lambda_1 = 0.2$. By the sample-based LP method, the optimal solution is $\hat{x}^{\mathrm{RA}_1} = 20.7824$. For the entropic exponential utility function model, defined as

$$\min_{x\geq 0}\frac{1}{\lambda_2}\ln\mathbb{E}\left[\mathrm{e}^{-\lambda_2\Pi(x;D)}\right]. \tag{2.12}$$

I set $\lambda_2 = 0.0072$, which results in a sample-based solution $\hat{x}^{\mathrm{RA}_2} = 20.7786$. For the mean-variance model, defined as

$$\min_{x\geq 0} -\mathbb{E}\left[\Pi(x;D)\right] + \lambda_3\mathbb{V}\mathrm{ar}\left[\Pi(x;D)\right]. \tag{2.13}$$

I set $\lambda_3 = 0.0037$, which results in a sample-based solution $\hat{x}^{\mathrm{RA}_3} = 20.7918$. Then I change the unit of r (price), c (cost) and s (salvage value) from dollar to 30 cents, 10 cents, 3 cents and 1 cent while keeping all other parameters unchanged. My results are summarized in Table 2.3.

As one can see from this table, while the numerical solution under a coherent measure of risk is invariant with respect to the unit system, it varies significantly under other risk measures.

2.4 Basic Analytical Results

In this section, I prove two fundamental results for a multi-product risk-averse newsvendor model. As I do not assume independent demands for the two results in this section, Propositions 1 and 2 hold true both in independent and dependent demands. These two Propositions also take a role of key intermediate steps for further analysis in Sects. 2.5–2.7.

Proposition 1 (Convexity of the Model). *If $\rho[\cdot]$ is a coherent measure of risk, then $\rho[\Pi(x,D)]$ is a convex function of x.*

Proof. I first note that $\Pi(x,D) = \sum_{j=1}^{n} \Pi_j(x_j,D_j)$ is concave in x. That is, for any $0 \le \alpha \le 1$ and all x and y,

$$\Pi(\alpha x + (1-\alpha)y, D) \ge \alpha\Pi(x,D) + (1-\alpha)\Pi(y,D) \quad \text{for all } D.$$

Using the monotonicity axiom, I obtain

$$\begin{aligned}\rho[\Pi(\alpha x + (1-\alpha)y, D)] &\le \rho[\alpha\Pi(x,D) + (1-\alpha)\Pi(y,D)] \\ &\le \alpha\rho[\Pi(x,D)] + (1-\alpha)\rho[\Pi(y,D)].\end{aligned}$$

The second inequality follows by the axiom of convexity. □

Proposition 1 shows the convexity of my model. It means the convexity preserves in a risk-averse model as well as in a risk-neutral model. Observe that I did not use the axiom of positive homogeneity, and thus Proposition 1 extends to more general models (e.g., convex measures of risk). I next prove the intuitively clear statement that identical products should be ordered in equal quantities under coherent measures of risk.

Proposition 2 (Symmetry of the Solution). *Assume that all products are identical, i.e., prices, ordering costs, and salvage values are the same across all products. Furthermore, let the joint probability distribution of the demand be symmetric, that is, invariant with respect to permutations of the demand vector. Then, for every law-invariant coherent measure of risk $\rho[\cdot]$, one of the optimal solutions of problem (2.4) is a vector with equal coordinates, $\hat{x}_1^{RA} = \hat{x}_2^{RA} = \cdots = \hat{x}_n^{RA}$.*

Proof. An optimal solution exists, because with no loss of generality I can assume that x is bounded by some large constant, and $\rho[\Pi(x,D)]$ is continuous with respect to x (see Ruszczyński and Shapiro 2006a).

Let me consider an arbitrary order vector $x = (x_1,\ldots,x_n)$ and let P be an $n \times n$ permutation matrix. Then, the distribution of profit associated with Px is the same as that associated with x. There are $n!$ different permutations of x and let me denote them $x^1,\ldots,x^{n!}$. Consider the point

$$y = \frac{1}{n!}\sum_{i=1}^{n!} x^i.$$

It has all coordinates equal to the average of the coordinates x_j. As the joint probability distribution of $D_1, D_2, \ldots, D_n$ is symmetric, the distribution of $\Pi(x^i, D)$ is the same for each i. By Proposition 1 and law invariance of $\rho[\cdot]$ I obtain

$$\rho[\Pi(y,D)] \leq \frac{1}{n!}\sum_{i=1}^{n!}\rho[\Pi(x^i,D)] = \rho[\Pi(x,D)].$$

This means that for every plan x, the corresponding plan y with equal orders is at least as good. As an optimal plan exists, there is an optimal plan with equal orders. □

Note that Proposition 2 only requires symmetric joint demand distribution, but not independent demands.

2.5 Analytical Results for Independent Demands

In this section, I assume demand independence and provide two analytical results (impact of degree of risk aversion and impact of shift in mean demand) for the multi-product newsvendor model under coherent risk measures. First, to study the impact of the degree of risk aversion, let me first focus on a specific variability measure—the weighted mean-deviation from quantile, given by (2.7). The corresponding measure of risk has the form,

$$\rho[V] = -\mathbb{E}[V] + \lambda r_\beta[V]. \tag{2.14}$$

By (2.8), I can write

$$\rho[V] = -(1-\lambda\beta)\mathbb{E}[V] + \lambda\beta\,\mathrm{AVaR}_\beta[V]. \tag{2.15}$$

I consider the problem

$$\min_{x\geq 0}\left\{-\mathbb{E}[\Pi(x,D)] + \lambda r_\beta[\Pi(x,D)]\right\}. \tag{2.16}$$

Proposition 3 (Monotonicity of the Solution with a mean-deviation from quantile). *Assume that all products are identical and demands for all products are iid (independently and identically distributed) and have a continuous distribution. Let $\hat{x}^{\mathrm{RA}_1}$ be the solution of problem (2.16) for $\lambda = \lambda_1 > 0$, having equal coordinates. If $\lambda_2 \geq \lambda_1$ then there exists a solution $\hat{x}^{\mathrm{RA}_2}$ of problem (2.16) for $\lambda = \lambda_2$, having equal coordinates and such that $\hat{x}_j^{\mathrm{RA}_2} \leq \hat{x}_j^{\mathrm{RA}_1}$, $j = 1,\ldots,n$.*

For the proof of Proposition 3, refer to Choi et al. (2011). Then, my goal is to extend the monotonicity property to all law-invariant coherent measures of risk. Observe that my assumption about continuous distribution of the demand implies that the probability space is nonatomic. Consider the problem

$$\min_{x\geq 0}\{-\mathbb{E}[\Pi(x,D)] + \lambda\varkappa_{\mathscr{M}}[\Pi(x,D)]\}, \tag{2.17}$$

where $\varkappa_{\mathscr{M}}[V]$ is given by (2.11).

Proposition 4 (Monotonicity of the Solution with every coherent measure of risk). *Assume that all products are identical and demands for all products are iid and have a continuous distribution. Let* $\hat{x}^{\mathrm{RA}_1}$ *be the solution of problem* (2.17) *for* $\lambda = \lambda_1 > 0$*, having equal coordinates. If* $\lambda_2 \geq \lambda_1$ *then there exists a solution* $\hat{x}^{\mathrm{RA}_2}$ *of problem* (2.17) *for* $\lambda = \lambda_2$*, having equal coordinates and such that* $\hat{x}_j^{\mathrm{RA}_2} \leq \hat{x}_j^{\mathrm{RA}_1}$*,* $j = 1, \ldots, n$*.*

Proof. As in the proof of Proposition 3, each function $x \mapsto r_\beta[\Pi(x,D)]$ is nondecreasing, for every $\beta \in (0,1)$. Then the integral over β with respect to any nonnegative measure μ is nondecreasing as well. Taking the supremum in (2.11) does not change this property. Therefore, Proposition 4 holds true also for the mean–risk model with the risk $r[\cdot] = \varkappa_{\mathscr{M}}[\cdot]$. □

Finally, I discuss the impact of the shift in mean demand on the optimal order quantities under general coherent measures of risk.

Proposition 5 (Impact of the Shift in Mean Demand). *Assume that all products are identical and demands for all products are iid except that* $\mu_j = \mathbb{E}[D_j]$*,* $j = 1, \ldots, n$*. If* $\mu_1 \geq \mu_2 \geq \cdots \geq \mu_n$*, then* $\hat{x}_1^{\mathrm{RA}} \geq \hat{x}_2^{\mathrm{RA}} \geq \cdots \geq \hat{x}_n^{\mathrm{RA}}$*.*

Proof. Consider the demand vector $\tilde{D}_j = D_j - \mu_j + \mu_1$. As it has identical and *iid* components, by Proposition 2 there exists an optimal order vector $\tilde{x}$ with equal coordinates: $\tilde{x}_1 = \tilde{x}_2 = \cdots = \tilde{x}_n$, for the risk-averse multi-product newsvendor with $\tilde{D}$ as the demand vector. I can interpret the demand D as a sum of the random demand $\tilde{D}$ and a deterministic demand vector h with coordinates $h_j = \mu_j - \mu_1$. If $\tilde{x}_j > 0$, then by the Translation Equivariance axiom, it is easy to see that $\hat{x} = \tilde{x} + h$ is the solution of the problem

$$\min_{x \geq 0} \rho[\Pi(x,D)],$$

for *every* coherent measure of risk $\rho[\cdot]$. □

2.6 Asymptotic Analysis and Closed-Form Approximations

2.6.1 Asymptotic Optimality of Risk-Neutral Solutions

In this section, I study the asymptotic behavior of the risk-averse newsvendor model when the number of products tends to infinity. I assume heterogenous products with independent demands.

I start from the derivation of error bounds for the risk-neutral solution. Consider a sequence of products $j = 1, 2, \ldots,$ with corresponding prices r_j, costs c_j, and salvage values s_j. I assume that $s_j < c_j < r_j$, and that all these quantities are uniformly bounded for $j = 1, 2 \ldots.$

Consider the risk-neutral optimal order quantities

$$\hat{x}_j^{\mathrm{RN}} = \bar{F}_j^{-1}\left(\frac{\bar{c}_j}{\bar{r}_j}\right), \quad j = 1,2,\dots. \tag{2.18}$$

I assume that the following conditions are satisfied:

(i) There exist $x^{\min} > 0$ and $x^{\max}$ such that

$$x^{\min} \le \hat{x}_j^{\mathrm{RN}} \le x^{\max}, \quad j = 1,2,\dots.$$

(ii) There exists $\sigma_{\min} > 0$ such that

$$\mathbb{V}\mathrm{ar}\left[\min\left(\hat{x}_j^{\mathrm{RN}}, D_j\right)\right] \ge \sigma_{\min}^2, \quad j = 1,2,\dots.$$

My intention is to evaluate the quality of the risk-neutral solution $\hat{x}^{\mathrm{RN}}$ in the risk-averse problem

$$\min_{x_1,\dots,x_n} \rho\left[\frac{1}{n}\sum_{j=1}^n \Pi_j(x_j, D_j)\right]. \tag{2.19}$$

Observe that in problem (2.19) I consider the average profit per product, rather than the total profit, as in problem (2.4). The reason is that I intend to analyze properties of the optimal value of this problem as $n \to \infty$ and I want the limit of the objective value of problem (2.19) to exist. Owing to the Positive Homogeneity axiom, problems (2.19) and (2.4) are equivalent.

I denote by $\hat{\rho}_n$ the optimal value of problem (2.19). I also introduce the following notation,

$$\mu_j^{\mathrm{RN}} = \mathbb{E}\left[\min(\hat{x}_j^{\mathrm{RN}}, D_j)\right], \qquad \bar{\mu}_n = \frac{1}{n}\sum_{j=1}^n \bar{r}_j \mu_j^{\mathrm{RN}},$$

$$\left(\sigma_j^{\mathrm{RN}}\right)^2 = \mathbb{V}\mathrm{ar}\left[\min\left(\hat{x}_j^{\mathrm{RN}}, D_j\right)\right], \qquad \bar{s}_n^2 = \frac{1}{n^2}\sum_{j=1}^n \bar{r}_j^2 \left(\sigma_j^{\mathrm{RN}}\right)^2.$$

Finally, I denote by $\mathscr{N}$ the standard normal variable. Then, I will show asymptotic convergence of risk-neutral solution to the true risk-averse solution.

Proposition 6 (Asymptotic Convergence of Risk-Neutral Solution with Error Bound). *Assume that $\rho[\cdot]$ is a law-invariant coherent measure of risk and the space $(\Omega, \mathscr{F}, P)$ is nonatomic. Then*

$$\rho\left[\frac{1}{n}\sum_{j=1}^n \Pi_j(\hat{x}_j^{\mathrm{RN}}, D_j)\right] \le \min_{x_1,\dots,x_n} \rho\left[\frac{1}{n}\sum_{j=1}^n \Pi_j(x_j, D_j)\right] + O\left(\frac{1}{\sqrt{n}}\right). \tag{2.20}$$

For the proof, refer to Choi et al. (2011). Asymptotically, the difference between the optimal objective value (the first term of the right-hand side of (2.20)) and the value obtained by using the risk-neutral solution (the term in the left-hand side of (2.20)) disappears at the rate of $1/\sqrt{n}$. Such difference can be considered as the error bound of a risk-neutral solution given as an "o" function of $1/\sqrt{n}$. Thus, for a firm dealing with very many products having independent demands, the risk-neutral solution is a reasonable sub-optimal alternative to the risk-averse solution.

2.6.2 Adjustments in the Mean-Deviation from Quantile Model

In this section, I develop close-form approximations to the optimal risk-averse solution when the number of products is moderately large. My idea is to use the risk-neutral solution as the starting point, and to calculate an appropriate correction to account for risk aversion.

I first consider the mean-deviation from quantile model in which the measure of variability is defined at (2.7). Recall that the corresponding mean–risk model in (2.14) is equivalent to the minimization of a combination of the mean and the Conditional Value-at-Risk, as in (2.15). I then consider the general coherent risk measure in Sect. 2.6.3. I finally discuss several iterative methods that are based on the approximations in Sect. 2.6.4.

I use the notation $Z_x^n = \frac{1}{n}\sum_{j=1}^n \bar{r}_j \min(x_j, D_j)$ (with x as a subscript to stress the dependence of Z_x^n on x). Using (2.1) and (2.2), I can calculate the average profit per product as follows:

$$\bar{\Pi}(x,D) = \frac{1}{n}\sum_{j=1}^n \Pi_j(x_j,D_j) = -\frac{1}{n}\sum_{j=1}^n \bar{c}_j x_j + Z_x^n.$$

Thus,

$$\begin{aligned}\rho[\bar{\Pi}(x,D)] &= \frac{1}{n}\sum_{j=1}^n \bar{c}_j x_j + \left(-\mathbb{E}[Z_x^n] + \lambda r_\beta(Z_x^n)\right)\\ &= \frac{1}{n}\sum_{j=1}^n \bar{c}_j x_j + \left(\mathbb{E}[Z_x^n](\lambda\beta - 1) - \lambda\beta \max_{\eta\in\mathbb{R}}\left\{\eta - \frac{1}{\beta}\mathbb{E}\left[(\eta - Z_x^n)^+\right]\right\}\right).\end{aligned} \tag{2.21}$$

Let me denote $\hat{\eta}$ to be the maximizer in (2.21), among $\eta \in \mathbb{R}$, at a fixed x. $\hat{\eta}$ is the β-quantile of Z_x^n. To take the partial derivative of $\rho[\bar{\Pi}(x,D)]$ with respect to x_j, I consider two cases.

Case (i): $\hat{\eta} < \frac{1}{n}\sum_{j=1}^n \bar{r}_j x_j$.

Assuming that the quantile $\hat{\eta}$ is unique and differentiating (2.21), I observe again that

$$\frac{\partial \rho[\bar{\Pi}(x,D)]}{\partial x_j} = \frac{\bar{c}_j}{n} + \frac{\bar{r}_j(\lambda\beta - 1)}{n}\mathbb{P}[D_j > x_j] - \frac{\bar{r}_j\lambda}{n}\mathbb{P}\left[\{Z_x^n < \hat{\eta}\} \cap \{D_j > x_j\}\right]. \tag{2.22}$$

Here I used (Bonnans and Shapiro 2000, Theorem 4.13) to avoid differentiating with respect to $\hat{\eta}$.

Let me analyze the last term on the right-hand side in (2.22) for $j = 1,2,\ldots,n$:

$$\begin{aligned}\mathbb{P}\left[\{Z_x^n < \hat{\eta}\} \cap \{D_j > x_j\}\right] &= \mathbb{P}\left[Z_x^n < \hat{\eta} | D_j > x_j\right]\mathbb{P}[D_j > x_j] \\ &= \mathbb{P}\left[\frac{1}{n}\sum_{k\neq j}^{n} \bar{r}_k \min(x_k, D_k) < \hat{\eta} - \frac{\bar{r}_j x_j}{n}\right] \cdot \mathbb{P}[D_j > x_j]. \end{aligned} \tag{2.23}$$

Suppose $x_j \geq x_{\min}$, $j = 1,2,\ldots,$. Owing to conditions (i) and (ii), exactly as in Sect. 2.6.1, for large n the random variable Z_x^n is approximately normally distributed with the mean $\bar{\mu}_n = \frac{1}{n}\sum_{j=1}^n \bar{r}_j\mu_j$ and the variance $\bar{s}_n^2 = \frac{1}{n^2}\sum_{j=1}^n \bar{r}_j^2\sigma_j^2$, where $\mu_j = \mathbb{E}[\min\{x_j, D_j\}]$ and $\sigma_j^2 = \mathbb{V}\mathrm{ar}(\min\{x_j, D_j\})$. Under normal approximation, the β-quantile of Z_x^n can be approximated by $\hat{\eta} \simeq \bar{\mu}_n + z_\beta \bar{s}_n$, where z_β is the β-quantile of the standard normal variable. Similarly, $\frac{1}{n-1}\sum_{k\neq j}^n \bar{r}_k \min(x_k, D_k)$ is approximately normal with mean $\frac{1}{n-1}\sum_{k\neq j}^n \bar{r}_k\mu_k$ and variance $\frac{1}{(n-1)^2}\sum_{k\neq j}^n \bar{r}_k^2\sigma_k^2$. Using these approximations and denoting by $\mathscr{N}$ the standard normal random variable, I obtain:

$$\begin{aligned}\mathbb{P}\left[\frac{1}{n}\sum_{k\neq j}^{n} \bar{r}_k \min(x_k, D_k) < \hat{\eta} - \frac{\bar{r}_j x_j}{n}\right] &\simeq \mathbb{P}\left[\mathscr{N} < \frac{-\bar{r}_j(x_j - \mu_j) + z_\beta\sqrt{\sum_{k=1}^n \bar{r}_k^2\sigma_k^2}}{\sqrt{\sum_{k\neq j}\bar{r}_k^2\sigma_k^2}}\right] \\ &= \mathbb{P}\left[\mathscr{N} < \frac{-\bar{r}_j(x_j - \mu_j)}{\sqrt{n-1}\gamma_{nj}} + z_\beta\sqrt{1 + \frac{\bar{r}_j^2\sigma_j^2}{(n-1)\gamma_{nj}^2}}\right], \end{aligned} \tag{2.24}$$

where $\gamma_{nj} = \sqrt{\frac{1}{n-1}\sum_{k\neq j}\bar{r}_k^2\sigma_k^2}$. As $\bar{r}_k^2\sigma_k^2$ is uniformly bounded from above and below across all products, I conclude that γ_{nj} is bounded from above and below for all j and n.

This estimate can be put into (2.23), and thus (2.22) can be approximated as follows:

$$\begin{aligned}\frac{\partial \rho[\bar{\Pi}(x,D)]}{\partial x_j} &\simeq \frac{\bar{c}_j}{n} + \frac{\bar{r}_j\mathbb{P}[D_j > x_j]}{n} \\ &\times \left(\lambda\beta - 1 - \lambda\mathbb{P}\left[\mathscr{N} < \frac{-\bar{r}_j(x_j - \mu_j)}{\sqrt{n-1}\gamma_{nj}} + z_\beta\sqrt{1 + \frac{\bar{r}_j^2\sigma_j^2}{(n-1)\gamma_{nj}^2}}\right]\right). \end{aligned} \tag{2.25}$$

My next step is to approximate the probability on the right-hand side of (2.25). To this end, I derive its limit and calculate a correction to this limit for a finite n. When $n \to \infty$, I have

$$\mathbb{P}\left[\mathcal{N} < \frac{-\bar{r}_j(x_j - \mu_j)}{\sqrt{n-1}\gamma_{nj}} + z_\beta \sqrt{1 + \frac{\bar{r}_j^2 \sigma_j^2}{(n-1)\gamma_{nj}^2}}\right] \to \beta \tag{2.26}$$

and thus

$$\frac{\partial \rho[\bar{\Pi}(x,D)]}{\partial x_j} \to \frac{1}{n}(\bar{c}_j - \bar{r}_j \mathbb{P}[D_j > x_j]).$$

This means that the conditions of the risk-averse solution

$$\frac{\partial \rho[\bar{\Pi}(x,D)]}{\partial x_j} = 0, \quad j = 1, 2, \ldots, n, \tag{2.27}$$

approaches that of the risk-neutral solution in (2.18). Thus the risk-neutral solution will be used as the base value, to which corrections will be calculated.

I can estimate the difference between the probability in (2.26) and β for a large but finite n, by assuming that x is close to $\hat{x}^{\mathrm{RN}}$. Thus, μ_j is close to $\mu_j^{\mathrm{RN}} = \mathbb{E}[\min\{\hat{x}_j^{\mathrm{RN}}, D_j\}]$ and σ_j is close to $\sigma_j^{\mathrm{RN}} = \sqrt{\mathbb{V}\mathrm{ar}(\min\{\hat{x}_j^{\mathrm{RN}}, D_j\})}$. Considering only the leading term with respect to $1/\sqrt{n-1}$, I obtain

$$\mathbb{P}\left[\mathcal{N} < \frac{-\bar{r}_j(x_j - \mu_j)}{\sqrt{n-1}\gamma_{nj}} + z_\beta \sqrt{1 + \frac{\bar{r}_j^2 \sigma_j^2}{(n-1)\gamma_{nj}^2}}\right] \simeq \mathbb{P}\left[\mathcal{N} < \frac{-\bar{r}_j(\hat{x}_j^{\mathrm{RN}} - \mu_j^{\mathrm{RN}})}{\sqrt{n-1}\gamma_{nj}^{\mathrm{RN}}} + z_\beta\right],$$

where $\gamma_{nj}^{\mathrm{RN}} = \sqrt{\frac{1}{n-1}\sum_{k \neq j} \bar{r}_k^2 (\sigma_k^{\mathrm{RN}})^2}$. The last probability can be estimated by the linear approximation derived at z_β. Observing that $P[\mathcal{N} < z_\beta] = \beta$ and that its derivative at $z = z_\beta$ is the standard normal density at z_β, I get

$$\mathbb{P}\left[\mathcal{N} < \frac{-\bar{r}_j(\hat{x}_j^{\mathrm{RN}} - \mu_j^{\mathrm{RN}})}{\sqrt{n-1}\gamma_{nj}^{\mathrm{RN}}} + z_\beta\right] \simeq \beta - \delta_{nj}^{\mathrm{RN}},$$

with

$$\delta_{nj}^{\mathrm{RN}} = \frac{\mathrm{e}^{-z_\beta^2/2}}{\sqrt{2\pi}} \frac{\bar{r}_j(\hat{x}_j^{\mathrm{RN}} - \mu_j^{\mathrm{RN}})}{\sqrt{n-1}\gamma_{nj}^{\mathrm{RN}}}, \quad j = 1, \ldots, n. \tag{2.28}$$

These estimates can be substituted to (2.25) for the derivative and yield

$$\frac{\partial \rho[\bar{\Pi}(x,D)]}{\partial x_j} \simeq \frac{\bar{c}_j}{n} + \frac{\bar{r}_j}{n}\left(-1 + \lambda \delta_{nj}^{\mathrm{RN}}\right)\mathbb{P}[D_j > x_j]. \tag{2.29}$$

Using the above approximations of the derivatives in (2.27), I obtain the first-order approximation of the risk-averse solution:

$$\hat{x}_j^{\mathrm{APR}} = \bar{F}_j^{-1}\left[\frac{\bar{c}_j}{\bar{r}_j(1-\delta_{nj}^{\mathrm{RN}}\lambda)}\right], \quad j=1,2,\ldots,n. \tag{2.30}$$

Clearly, this approximation of $\hat{x}_j^{\mathrm{APR}}$ is increasing in n, decreasing in λ, and tends to the risk-neutral solution as $n\to\infty$. Similar to the analysis in Sect. 2.6.1, the error bound of this approximation in (2.30) is given as follows:

$$0 \le -\frac{\partial\rho\left(\bar{\Pi}\left(\hat{x}_j^{\mathrm{APR}},D\right)\right)}{\partial x_j} \le O(1/n^{3/2}). \tag{2.31}$$

It implies that as the number of products increases, the convergence rate of my approximate solution to the risk-averse solution in (2.31), $O(1/n^{3/2})$, is much faster than the rate of the risk-neutral solution in (2.20), $O(1/n^{1/2})$.

Case (ii): $\hat{\eta} = \frac{1}{n}\sum_{j=1}^{n}\bar{r}_j x_j$.

I have

$$\rho[\bar{\Pi}(x,D)] = \frac{1}{n}\sum_{j=1}^{n}\bar{c}_j x_j + \Bigg(\mathbb{E}[Z_x^n](\lambda\beta-1) \\ -\lambda\beta\left\{\frac{1}{n}\sum_{j=1}^{n}\bar{r}_j x_j - \frac{1}{\beta}\mathbb{E}\left[\frac{1}{n}\sum_{j=1}^{n}\bar{r}_j x_j - Z_x^n\right]\right\}\Bigg).$$

Taking derivative with respect to x_j yields,

$$\frac{\partial\rho[\bar{\Pi}(x,D)]}{\partial x_j} = \frac{1}{n}\left[\bar{c}_j + \bar{r}_j\lambda(1-\beta) + \bar{r}_j\mathbb{P}[D_j > x_j]\left(\lambda(\beta-1)-1\right)\right].$$

Equating the right-hand side to 0, I get

$$\hat{x}_j^{\mathrm{RA}} = \bar{F}_j^{-1}\left(\frac{\bar{c}_j + \bar{r}_j\lambda(1-\beta)}{\bar{r}_j(1+\lambda(1-\beta))}\right). \tag{2.32}$$

Note that the solution in Case (ii) is an exact solution and free of the number of products, n. Clearly, if $\lambda=0$, $\hat{x}_j^{\mathrm{RA}} = \hat{x}_j^{\mathrm{RN}}$. As λ increases, $\hat{x}_j^{\mathrm{RA}}$ is decreasing. For any $0\le\lambda\le 1/\beta$, $\hat{x}_j^{\mathrm{RA}}$ is well defined.

It should be emphasized that Case (i) is more important, because for large n the distribution of Z_x^n is close to normal and for a small β, the β-quantile of Z_x^n tends to be smaller than $\frac{1}{n}\sum_{j=1}^{n}\bar{r}_j x_j$, for the values of x of interest.

Consider the special case of identical products. With a slight abuse of notation, let $c_j = c$, $r_j = r$ and $s_j = s$ for all $j = 1, 2, \ldots, n$. In Case (i), the first-order approximation of the risk-averse solution yields:

$$\frac{\mathrm{d}\rho[\bar{\Pi}(x,D)]}{\mathrm{d}x} \simeq \bar{c} + \bar{r}\mathbb{P}[D_1 > x](\delta_n^{\mathrm{RN}}\lambda - 1),$$

with

$$\delta_n^{\mathrm{RN}} = \frac{\mathrm{e}^{-z_\beta^2/2}}{\sqrt{2\pi}} \frac{\hat{x}^{\mathrm{RN}} - \mu_x^{\mathrm{RN}}}{\sqrt{n-1}\sigma_x^{\mathrm{RN}}}, \tag{2.33}$$

where $\hat{x}^{\mathrm{RN}}$, μ_x^{RN}, and σ_x^{RN} are the counterparts of $\hat{x}_j^{\mathrm{RN}}$, μ_j^{RN}, and σ_j^{RN}, respectively. Equating the right hand side to 0, I obtain

$$\hat{x}_1^{\mathrm{APR}} = \bar{F}_1^{-1}\left(\frac{\bar{c}}{\bar{r}(1 - \delta_n^{\mathrm{RN}}\lambda)}\right), \quad j = 1, \ldots, n. \tag{2.34}$$

Here, (2.34) is similar to (2.30) except that the terms $\bar{c}_j$, $\bar{r}_j$ and $\delta_{nj}^{\mathrm{RN}}$ are now identical for all j. In Case (ii), (2.32) reduces to

$$\hat{x}_1^{\mathrm{RA}} = \bar{F}_1^{-1}\left(\frac{\bar{c} + \bar{r}\lambda(1-\beta)}{\bar{r}(1 + \lambda(1-\beta))}\right).$$

In the special case of a single-product problem, by (2.22) in Case (i) I obtain

$$\frac{\mathrm{d}\rho[\bar{\Pi}(x,D)]}{\mathrm{d}x} = \bar{c} + \bar{r}(\lambda\beta - 1)\mathbb{P}[D > x] - \bar{r}\lambda\mathbb{P}[\{Z_x < \hat{\eta}\} \cap \{D > x\}],$$

where $Z_x = \min(x, D)$. Observe that in Case (i),$\mathbb{P}[\{Z_x < \hat{\eta}\} \cap \{D > x\}] = \mathbb{P}[Z_x < \hat{\eta} | D > x]\,\mathbb{P}[D > x] = 0$. Therefore, $\frac{\mathrm{d}\rho[\bar{\Pi}(x,D)]}{\mathrm{d}x} = \bar{c} + \bar{r}(\lambda\beta - 1)\mathbb{P}[D > x]$. This yields the exact solution of the single product problem

$$\hat{x}^{\mathrm{RA}} = \bar{F}^{-1}\left(\frac{\bar{c}}{\bar{r}(1 - \lambda\beta)}\right) \leq \bar{F}^{-1}\left(\frac{\bar{c}}{\bar{r}}\right) = \hat{x}^{\mathrm{RN}}.$$

This special case solution is the same as the solution obtained by Gotoh and Takano (2007). To determine whether Case (i) or Case (ii) applies, one can compute $\hat{x}^{\mathrm{RA}}$ for both cases, and then compute $\hat{\eta}$ to check the case conditions.

2.6.3 General Law-Invariant Coherent Measures of Risk

So far my analysis focused on a special risk measure, weighted mean-deviation from quantile, given in (2.7). I now generalize the results to any law-invariant coherent risk measure $\rho[\cdot]$.

Consider problem (2.17) where $\varkappa_{\mathscr{M}}[V]$ is given by (2.11). By Kusuoka theorem, for nonatomic spaces, every law-invariant coherent measure of risk has such representation. Thus, I focus on Case (i) solution only. Then (2.21) can be replaced by

$$\rho[\bar{\Pi}(x,D)] = \frac{1}{n}\sum_{j=1}^{n} \bar{c}_j x_j + \sup_{\mu\in\mathscr{M}} \int_0^1 \Big(\mathbb{E}[Z_x^n](\lambda\beta - 1) \\ -\lambda\beta \max_{\eta\in\mathbb{R}} \Big\{ \eta - \frac{1}{\beta}\mathbb{E}\left[(\eta - Z_x^n)^+\right] \Big\} \Big)\, \mu(\mathrm{d}\beta).$$

Suppose the maximum over $\mathscr{M}$ is attained at a unique measure $\hat{\mu}$ (this is certainly true for *spectral* measures of risk, where the set $\mathscr{M}$ has just one element). Similarly to (2.29),

$$\frac{\partial\rho[\bar{\Pi}(x,D)]}{\partial x_j} \simeq \frac{\bar{c}_j}{n} + \frac{\bar{r}_j}{n}\left(-1 + \lambda\int_0^1 \delta_{nj}^{\mathrm{RN}}(\beta)\,\hat{\mu}(\mathrm{d}\beta)\right)\mathbb{P}[D_j > x_j]. \tag{2.35}$$

I denote here the quantity given in (2.28) by $\delta_{nj}^{\mathrm{RN}}(\beta)$, to stress its dependence on β. Let me approximate $\hat{\mu}$ by the measure $\hat{\mu}^{\mathrm{RN}}$, obtained for the risk-neutral solution $\hat{x}^{\mathrm{RN}}$. Equating the approximate derivatives in (2.35) to zero, I obtain an approximate solution:

$$\hat{x}_j^{\mathrm{APR}} = \bar{F}_j^{-1}\left(\frac{\bar{c}_j}{\bar{r}_j\left(1 - \lambda\int_0^1 \delta_{nj}^{\mathrm{RN}}(\beta)\,\hat{\mu}^{\mathrm{RN}}(\mathrm{d}\beta)\right)} \right), \quad j = 1,2,\ldots,n. \tag{2.36}$$

Again, $\delta_{nj}^{\mathrm{RN}}(\beta) \downarrow 0$ as $n \to \infty$, and thus $\hat{x}_j^{\mathrm{APR}}$ increases in n and approaches the risk-neutral solution $\hat{x}_j^{\mathrm{RN}}$. This is consistent with Proposition 6 and the analysis in Sect. 2.6.2.

In the special case of identical products, the approximate solution is

$$\hat{x}_1^{\mathrm{APR}} = \bar{F}_1^{-1}\left(\frac{\bar{c}_j}{\bar{r}_j\left(1 - \lambda\int_0^1 \delta_{n}^{\mathrm{RN}}(\beta)\,\hat{\mu}^{\mathrm{RN}}(\mathrm{d}\beta)\right)} \right), \quad j = 1,2,\ldots,n, \tag{2.37}$$

where δ_n^{RN} is defined at (2.33).

In the single-product problem, I obtain

$$\rho[\bar{\Pi}(x,D)] = \bar{c}x + \sup_{\mu\in\mathscr{M}} \int_0^1 \Big(\mathbb{E}[Z_x](\lambda\beta - 1) - \lambda\beta \max_{\eta\in\mathbb{R}} \left\{ \eta - \frac{1}{\beta}\mathbb{E}\left[(\eta - Z_x)^+\right] \right\} \Big) \mu(\mathrm{d}\beta). \tag{2.38}$$

Assuming that $\hat{\mu}$ is the unique maximizer in (2.38), I obtain

$$\frac{\mathrm{d}\rho[\bar{\Pi}(x,D)]}{\mathrm{d}x} = \bar{c} + \int_0^1 (\bar{r}(\lambda\beta - 1)\mathbb{P}[D > x] - \bar{r}\lambda\mathbb{P}[\{Z_x < \hat{\eta}\} \cap \{D > x\}])\, \hat{\mu}(\mathrm{d}\beta).$$

Similarly to the model with mean-deviation from quantile case, $\mathbb{P}[\{Z_x < \hat{\eta}\} \cap \{D > x\}] = \mathbb{P}[Z_x < \hat{\eta} | D > x]\,\mathbb{P}[D > x] = 0$. Thus,

$$\begin{aligned}\frac{\mathrm{d}\rho[\bar{\Pi}(x,D)]}{\mathrm{d}x} &= \bar{c} + \int_0^1 (\bar{r}(\lambda\beta - 1)\mathbb{P}[D > x])\, \hat{\mu}(\mathrm{d}\beta) \\ &= \bar{c} + \bar{r}\left(-1 + \lambda \sup_{\mu\in\mathscr{M}} \int_0^1 \beta\hat{\mu}(\mathrm{d}\beta) \right) \mathbb{P}[D > x].\end{aligned}$$

Therefore, the closed-form exact solution for general coherent measures of risk is given by:

$$\hat{x}^{\mathrm{RA}} = \bar{F}^{-1}\left(\frac{\bar{c}}{\bar{r}(1-\lambda\bar{\beta})}\right) \le \bar{F}^{-1}\left(\frac{\bar{c}}{\bar{r}}\right) = \hat{x}^{\mathrm{RN}}, \quad \text{where } \bar{\beta} = \int_0^1 \beta\hat{\mu}^{\mathrm{RN}}(\mathrm{d}\beta).$$

2.6.4 Iterative Methods

So far, I discussed approximations based on expansions about the risk-neutral solution $\hat{x}^{\mathrm{RN}}$. But exactly the same argument can be used to develop an iterative method, in which the best approximation known so far is substituted for the risk-neutral solution. I explain the simplest idea for the approximation developed in Sect. 2.6.2; the same idea applies to general coherent measures of risk discussed in Sect. 2.6.3.

The idea of the *iterative method* is to generate a sequence of approximations $\hat{x}^{(\nu)}$, $\nu = 0, 1, 2, \ldots$. I set $\hat{x}^{(0)} = \hat{x}^{\mathrm{RN}}$. Then I calculate $\hat{x}^{(1)}$ by applying (2.30). In the iteration $\nu = 1, 2, \ldots$, I use $\hat{x}^{(\nu)}$ instead of $\hat{x}^{\mathrm{RN}}$ in my approximation, calculating:

$$\mu_j^{(\nu)} = \mathbb{E}\left[\min\left\{\hat{x}_j^{(\nu)}, D_j\right\}\right], \quad \sigma_j^{(\nu)} = \sqrt{\mathbb{V}\mathrm{ar}\left(\min\left\{\hat{x}_j^{(\nu)}, D_j\right\}\right)},$$

$$\gamma_{nj}^{(\nu)} = \sqrt{\frac{1}{n-1}\sum_{k\neq j}\bar{r}_k^2\left\{\left(\sigma_k^{(\nu)}\right)\right\}^2}, \quad \delta_{nj}^{(\nu)} = \frac{\mathrm{e}^{-z_\beta^2/2}}{\sqrt{2\pi}}\frac{\bar{r}_j\left(\hat{x}_j^{(\nu)} - \mu_j^{(\nu)}\right)}{\sqrt{n-1}\gamma_{nj}^{(\nu)}}, \quad j = 1,\ldots,n.$$

Finally, (2.30) is applied to generate the next approximate solution $\hat{x}^{(\nu+1)}$, and the iteration continues.

The iterative method is efficient if the initial approximation $\hat{x}^{(0)}$ is sufficiently close to the risk-averse solution. This is true when the risk aversion coefficient $\kappa = \lambda\beta$ is close to zero or the number of products is very large. I must point out that the iterative method does not guarantee convergence to the optimal risk-averse solution. One reason is that my approximation in (2.30) may result in infeasible solutions as the term $\frac{\bar{c}_j}{\bar{r}_j\left(1-\delta_{nj}^{(\nu)}\lambda\right)}$ can be negative or greater than 1 (due to approximation). When this occurs less likely, one can say that the approximation is more stable. Generally, the approximation is more stable for larger number of products and smaller κ. To improve stability, I propose a more accurate method called the *continuation method*. In this approach, I apply the iterative method for a small value of κ, starting from the risk-neutral solution. Then I increase κ a little, and I apply the iterative method again, but starting from the best solution found for the previous value of κ. In this way, I gradually increase κ, until I reach the risk aversion coefficients which are of interest (usually, between 0 and 1). The stability of the iterative and continuation methods is summarized in Sect. 2.8.2.

2.7 Impact of Dependent Demands

In this section, I provide some insights on the impact of dependent demands. Due to significant analytical challenges, I focus on a two-product system and the mean-deviation from quantile model.

Under the risk-neutral measure, dependence of product demands has no impact on the optimal order quantities. However, under risk-averse measures, it can greatly affect the optimal order decisions for the newsvendor. Intuitively, positively (negatively) dependent demands entail larger (smaller) variability and thus increase (decrease) risk, as compared to independent demands. Thus, one tends to decrease (increase) the order quantity in case of positively (negatively) dependent demands relative to the case of independent demand.

To characterize the impact of demand dependence on the optimal order quantity under the coherent risk measure, I utilize the concept of "associated" random variables. Consider random variables $D_1, D_2, \ldots, D_n$, denote vector $D = (D_1, D_2, \ldots, D_n)$. The following definition is due to Esary et al. (1976); see Tong (1980) for a review.

Definition 1. *The random variables* $D_1, D_2, \ldots, D_n$ *are associated, if* $\mathbb{C}\text{ov}[f(D), g(D)] \geq 0$*, or, equivalently,* $\mathbb{E}[f(D)g(D)] \geq \mathbb{E}[f(D)]\mathbb{E}[g(D)]$*, for all nondecreasing real functions* f, g *for which* $\mathbb{E}[f(D)], \mathbb{E}[g(D)]$ *and* $\mathbb{E}[f(D)g(D)]$ *exist.*

Lemma 1. *(i) Any subset of a set of associated random variables is associated.*
(ii) If two sets of associated random variables are independent of each other, their union is a set of associated random variables.
(iii) Nondecreasing (or nonincreasing) functions of associated random variables are associated.
(iv) If $D_1, D_2, \ldots, D_n$ *are associated, then for all* $(y_1, y_2, \ldots, y_n) \in R^n$

$$\mathbb{P}\{D_1 \leq y_1, D_2 \leq y_2, \ldots, D_n \leq y_n\} \geq \Pi_{k=1}^n \mathbb{P}\{D_k \leq y_k\},$$
$$\mathbb{P}\{D_1 \geq y_1, D_2 \geq y_2, \ldots, D_n \geq y_n\} \geq \Pi_{k=1}^n \mathbb{P}\{D_k \geq y_k\}.$$

I refer to Tong (1980) for proofs.

Association is closely related to correlation. By (Tong 1980, p. 99), a set of multivariate normal random variables is associated if their correlation matrix has the structure l (Tong 1980, p. 13) in which the correlation coefficient $\rho_{ij} = \gamma_i \gamma_j$ for all $i \neq j$ and $0 \leq \gamma_i < 1$ for all i. This means that I can represent the demands as having one common factor:

$$D_i = \gamma_i D_0 + \Delta_i, \quad i = 1, \ldots, n,$$

where D_0 and Δ_i, $i = 1, \ldots, n$, are independent. A special case is the bi-variate normal random variable with a positive correlation coefficient.

Consider a system with two identical products and a solution with equal coordinates. Let $Z_x = \min\{x, D_1\} + \min\{x, D_2\}$. Clearly, $\Pi(x, D) = -2\bar{c}x + \bar{r}Z_x$ and

$$\rho(\Pi(x, D)) = 2\bar{c}x + \bar{r}\rho(Z_x), \tag{2.39}$$

$$\rho(Z_x) = \mathbb{E}(Z_x)(\lambda\beta - 1) - \lambda\beta \max_{\eta \in \mathbb{R}} \left\{\eta - \tfrac{1}{\beta}\mathbb{E}[(\eta - Z_x)^+]\right\}. \tag{2.40}$$

Let $\hat{\eta}$ be the maximizer. If $\hat{\eta}$ is not an atom of the distribution of Z_x, similar to Case (i) analysis in Sect. 2.6.2, I obtain

$$\frac{\mathrm{d}\rho(Z_x)}{\mathrm{d}x} = \frac{\mathrm{d}\mathbb{E}[Z_x]}{\mathrm{d}x}(\lambda\beta - 1) + \lambda \frac{\mathrm{d}\mathbb{E}[(\hat{\eta} - Z_x)^+]}{\mathrm{d}x},$$

where $\hat{\eta}$ is the β-quantile of Z_x and $\hat{\eta} < 2x$. Because the first term depends only on the marginal distributions of the demands, I focus on the second term, which is affected by the dependence of D_1 and D_2. I have

$$\frac{\mathrm{d}\mathbb{E}[(\hat{\eta} - Z_x)^+]}{\mathrm{d}x} = -\sum_{j=1}^{2} \mathbb{P}[\{Z_x < \hat{\eta}\} \cap \{D_j > x\}] = -2\mathbb{P}[\min\{x, D_2\} < \hat{\eta} - x, D_1 > x]. \tag{2.41}$$

Consider three cases of (D_1, D_2), with the same the marginal distributions of D_1 and D_2. In case 1, (D_1, D_2) are associated random variables, and I use $\hat{\eta}_P$ to denote the β-quantile of the corresponding Z_x; In case 2, (D_1, D_2) are independent with $\hat{\eta}_I$ as the β-quantile of Z_x; In case 3, $(D_1, -D_2)$ are associated random variables with $\hat{\eta}_N$ as the β-quantile of Z_x. I also let x_P^*, x_I^*, and x_N^* be the optimal order quantities in cases 1, 2, and 3, respectively.

Proposition 7 (Impact of Demand Correlation). *If* $\hat{\eta}_P \leq \hat{\eta}_I \leq \hat{\eta}_N < 2x$*, then*

$$x_P^* \leq x_I^* \leq x_N^*. \tag{2.42}$$

That is, positively (negatively) dependent (D_1, D_2) *results in smaller (larger) optimal order quantities than independent* (D_1, D_2)*.*

Proof. I first consider associated (D_1, D_2). I have

$$\begin{aligned}
&\mathbb{P}[\min\{x, D_2\} < \hat{\eta}_P - x, D_1 > x] \\
&\quad = \mathbb{P}[D_2 < \hat{\eta}_P - x, D_1 > x] \\
&\quad = \mathbb{P}[D_1 > x] - \mathbb{P}[D_2 \geq \hat{\eta}_P - x, D_1 > x] \leq \mathbb{P}[D_1 > x] - \mathbb{P}[D_2 \geq \hat{\eta}_P - x]\mathbb{P}[D_1 > x] \\
&\quad = \mathbb{P}[D_2 < \hat{\eta}_P - x]\mathbb{P}[D_1 > x] \leq \mathbb{P}[D_2 < \hat{\eta}_I - x]\mathbb{P}[D_1 > x].
\end{aligned}$$

The first inequality follows by Lemma 1 part (iv). The second inequality follows by $\hat{\eta}_P \leq \hat{\eta}_I$. Note that the last term corresponds to independent (D_1, D_2). Thus, by (2.41), associated (D_1, D_2) have the derivatives $\mathrm{d}\rho(Z_x)/\mathrm{d}x$ at least as large as independent (D_1, D_2), which implies that $x_P^* \leq x_I^*$.

I then consider associated $(D_1, -D_2)$. I obtain

$$\begin{aligned}
\mathbb{P}[D_2 < \hat{\eta}_N - x, D_1 > x] &= \mathbb{P}[-D_2 > -\hat{\eta}_N + x, D_1 > x] \\
&\geq \mathbb{P}[-D_2 > -\hat{\eta}_N + x]\mathbb{P}[D_1 > x] \\
&= \mathbb{P}[D_2 < \hat{\eta}_N - x]\mathbb{P}[D_1 > x] \geq \mathbb{P}[D_2 < \hat{\eta}_I - x]\mathbb{P}[D_1 > x].
\end{aligned}$$

The first inequality follows by Lemma 1 part (iv). The second inequality follows by $\hat{\eta}_I \leq \hat{\eta}_N$. Note that the last term corresponds to independent (D_1, D_2). Thus, by (2.41), associated $(D_1, -D_2)$ have the derivatives $\mathrm{d}\rho(Z_x)/\mathrm{d}x$ no larger than independent (D_1, D_2), which implies that $x_I^* \leq x_N^*$. □

The condition $\hat{\eta}_P \leq \hat{\eta}_I \leq \hat{\eta}_N$ holds when $Y_1 = \min\{x, D_1\}$ and $Y_2 = \min\{x, D_2\}$ follow bivariate normal distribution and $\beta \leq 0.5$. One can approximate the joint distribution of Y_1 and Y_2 very closely by bivariate normal when (D_1, D_2) follow bivariate normal and x is set to cover most of the demand, which is very likely in practice when the underage cost $r - c$ is much greater than the overage cost $c - s$.

2.8 Numerical Examples

The objective of this section is twofold. First, I study the accuracy and the convergence rates of the approximations. Second, I provide insights (in addition to the analysis in Sects. 2.5–2.7) on the impact of demand dependence and risk aversion. I first introduce the sample-based optimization method.

2.8.1 Sample-Based Optimization

In all examples considered, I apply sample-based optimization to solve the resulting stochastic programming problems. I generate a sample $D^1, D^2, \ldots, D^T$ of the demand vector, where

$$D^t = (d_{1t}, d_{2t}, \ldots, d_{nt}), \quad t = 1, \ldots, T.$$

Then I replace the original demand distribution by the empirical distribution based on the sample, that is, I assign to each of the sample points the probability $p_t = 1/T$. It is known that when $T \to \infty$, the optimal value of the sample problem approaches the optimal value of the original problem (see Shapiro 2007). In all my examples, I used $T = 10,000$.

For the empirical distribution, the corresponding optimization problem (2.16) has an equivalent linear programming formulation. For each $j = 1, \ldots, n$ and $t = 1, \ldots, T$, I introduce the variable w_{jt} to represent the salvaged number of product j in scenario t. The variable u_t represents the shortfall of the profit in scenario t to the quantile η. It is also convenient to introduce the parameter $\kappa = \lambda\beta$ to represent the relative risk aversion ($0 \le \kappa \le 1$). I obtain the formulation

$$\max \quad (1-\kappa)\sum_{j=1}^{n}\left[(r_j - c_j)x_j - (r_j - s_j)\sum_{t=1}^{T} p_t w_{jt}\right] + \kappa\left(\eta - \frac{1}{\beta}\sum_{t=1}^{T} p_t u_t\right) \tag{2.43}$$

$$\text{subject to} \quad \sum_{j=1}^{n}\left[(r_j - c_j)x_j - (r_j - s_j)w_{jt}\right] + u_t \ge \eta, \quad t = 1, \ldots, T,$$

$$x_j - d_{jt} \le w_{jt}, \quad j = 1, \ldots, n; \quad t = 1, \ldots, T,$$

$$w_{jt} \ge 0, \quad j = 1, \ldots, n; \quad t = 1, \ldots, T,$$

$$u_t \ge 0, \quad t = 1, \ldots, T,$$

$$x_j \ge 0, \quad j = 1, \ldots, n.$$

To explain this formulation, suppose the order quantities x_j are fixed. Then $w_{jt} = (x_j - d_{jt})^+$ and $u_t = (\eta - \Pi(x, D^t))^+$ are optimal, and I maximize with respect to η the last term in problem (2.43), that is,

$$\max_{\eta} \left\{ \eta - \frac{1}{\beta} \mathbb{E}\left[(\eta - \Pi(x,D))^+ \right] \right\} = -\mathrm{AVaR}_{\beta}\left[\Pi(x,D)\right].$$

In the last expression, I used (2.8). Therefore, (2.43) is equal to $(1-\kappa)\mathbb{E}\left[\Pi(x,D)\right] - \kappa\mathrm{AVaR}_{\beta}\left[\Pi(x,D)\right]$.

2.8.2 Accuracy of Approximations

In this section, I assess the accuracy of the closed-form approximations of Sect. 2.6. I first consider identical products, then nonidentical products.

For identical products, I assume that all products have identical cost structure, and *iid* demands. I set $r = 15$, $c = 10$, and $s = 7$. I set the demand distribution of each product to be lognormal with $\mu = 3$ and $\sigma = 0.4724$ (to achieve the desirable coefficient of variance (cv) of 0.5). Thus, the mean and standard deviation of each demand are $\mathrm{e}^{\mu+\sigma^2/2} = 22.46$ and $\mathrm{e}^{\mu+\sigma^2/2} \cdot \sqrt{(\mathrm{e}^{\sigma^2} - 1)} = 11.23$. Because the joint demand distribution is invariant with respect to the permutations of the demand vector, there exists an order vector with equal coordinates, which is optimal for the model.

I choose the number of products, n, to be 1, 3, 10, and 30, and I study the impact of the number of products on the gap between the sample-based LP solutions and the approximate solutions (generated by the iterative method with $\nu = 3$, see Sect. 2.6.4). The sample-based LP solutions can take hours to solve, especially for large n and T. For instance, with $n = 30$ and a sample size of 10,000, the running time by CPLEX 9.0 at an Intel Pentium 4 PC is 32,607 s for identical products and 50,889 s for heterogenous products. In contrast, the approximate solution can be obtained within one or two seconds. I use $\beta = 0.5$, that is, I am concerned with the shortfall below the median.

In my numerical study of identical products, I set the optimal order quantities for different products to be identical by Proposition 2. In model (2.43), all variables x_j are replaced by a single variable x. The corresponding results are illustrated in Fig. 2.1, where on the horizontal axis I display the relative risk aversion parameter $\kappa = \lambda\beta$. The term "exact," "numerical," and "approximation" represent the solution obtained by the exact calculation, the sample-based LP, and the closed-form approximation, respectively.

Figure 2.1 shows that my analytical solution is very close to the numerical solution when $n = 1$. This is obvious as my solution is exact for the single-product case (here, the case $\hat{\eta} = x$ is valid). In the case of a three-product model, the approximation does not work well, which is quite understandable as

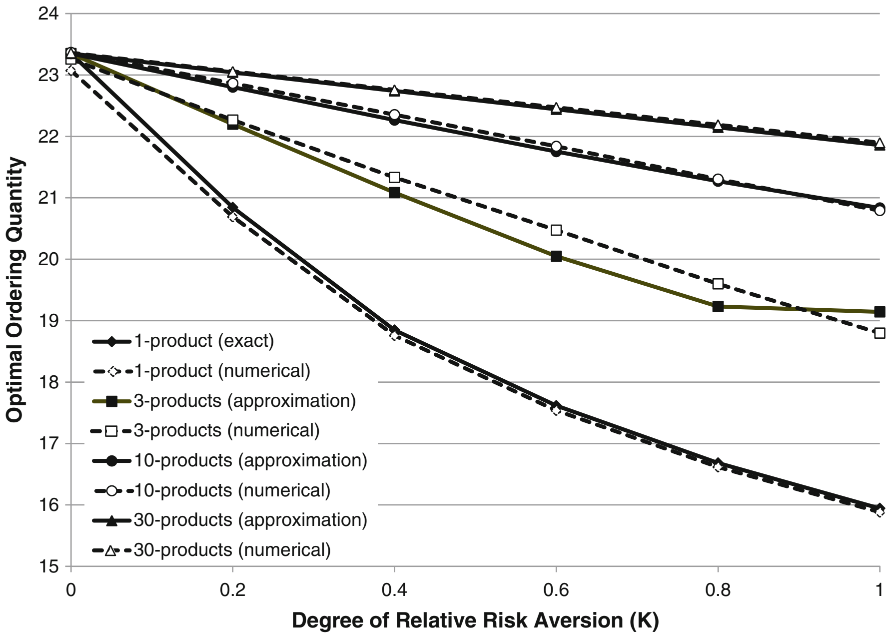


Fig. 2.1 Identical products with independent demands—Approximate or exact solutions *vs.* sample-based solutions. The terms "exact," "numerical," and "approximation" refer to exact solutions, solutions of the sample-based model, and closed-form approximations, respectively

the approximation is based on the Central Limit Theorem. As the number of products increases, my approximations become more accurate and the gap becomes negligible when $n \geq 10$. I also observe that the order quantities decrease as the degree of risk-aversion increases, which confirms Proposition 3; and as the number of products increases, the error of the risk neutral solution decreases (consistent with Proposition 6).

For independent but heterogenous products, I tested the accuracy of the approximations on 30 randomly generated problems, 10 for each number of products $n = 3, 10, 30$. At each value of $\kappa = 0.2, 0.4, 0.6, 0.8, 1$, I calculated the sample-based LP solution and an approximate solution by the continuation method with $\nu = 1$. My numerical study shows that the continuation method is much more stable and accurate than the iterative method with $\nu = 1$, especially for smaller numbers of products, when the difference between risk-neutral solution and risk-averse solution is larger (e.g., κ is larger). For $n = 30$, both methods work very well.

For each instance in which the continuation method can generate a feasible solution, I compute the absolute percentage error of the approximate solution relative to the sample-based LP solution, which is defined by the absolute difference between the approximate solution and the sample-based LP solution over the sample-based LP solution. For comparison, I also compute the absolute percentage

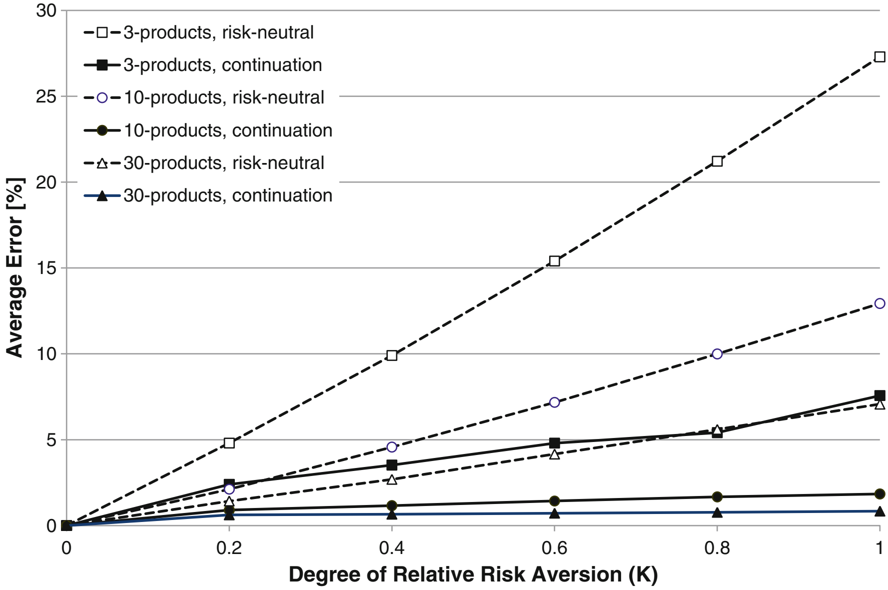


Fig. 2.2 Heterogeneous products with independent demands—The average percentage error of the approximate solutions and risk-neutral solutions

error of the risk-neutral solution relative to the sample-based LP solution. Then for each value of n and κ, I compute the average and maximum percentage error over all the solutions generated. The average (and maximum) percentage errors of the risk-neutral solutions and of the solutions obtained by the continuation method are displayed in Figs. 2.2 and 2.3, respectively).

In all cases, in terms of the average and maximum errors, my approximation outperforms the risk-neutral solution. Furthermore, in most cases, the improvement brought by the approximation is significant. Often, the approximation cuts the error of the risk-neutral solution by 3–6 times, although only one step of the continuation method was made at each κ. Second, I observe that the approximation is quite accurate for all cases of $n = 10$ and $n = 30$. However, the approximation does not work well for $n = 3$, which is similar to what I observed in the identical products case. Finally, I observe that the average and maximum errors of the risk-neutral solutions are decreasing in n, as established in Proposition 6.

2.8.3 Impact of Dependent Demands Under Risk Aversion

I first consider a simple system with two identical products, then a system with two heterogenous products. The numerical results here are obtained by the sample-based LP.

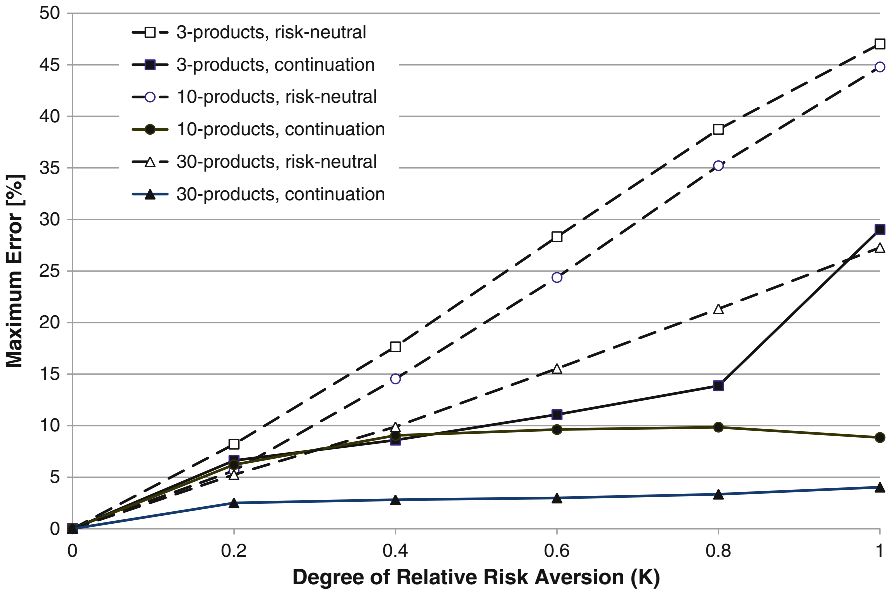


Fig. 2.3 Heterogeneous products with independent demands—The maximum percentage error of the approximate solutions and risk-neutral solutions

I choose the following cost parameters for the system with two identical products: $r_1 = r_2 = 15$, $c_1 = c_2 = 10$ and $s_1 = s_2 = 7$. I assume that demand follows bivariate lognormal distribution, which is generated by exponentiating a bivariate normal with the parameters $\mu_1 = \mu_2 = 3$, $\sigma_1 = \sigma_2 = 0.4724$ and a correlation coefficient of $-1, -0.8, -0.6, \ldots, 1$. Thus, the mean and standard deviation of each marginal distribution are 22.46 and 11.23 respectively with cv $= 0.5$. The numerical results are summarized in Fig. 2.4.

Consistent with my analysis in Sect. 2.5, risk aversion reduces optimal order quantities for independent or positively correlated demands, relative to the risk-neutral solution. But interestingly, this observation may not hold for strongly negatively correlated demands, where increased risk aversion can result in a greater optimal order quantity. To explain the intuition behind these counterexamples, let me consider two identical products with perfectly negatively correlated demands, D_1 and D_2. A larger order quantity, Q, increases negative correlation between the sales $\min(D_1, Q)$ and $\min(D_2, Q)$, and thus leads to smaller variability of the total sales $\min(D_1, Q) + \min(D_2, Q)$. Choi (2009) also studied a special case of a two-identical product system with bivariate uniform distribution and perfectly negative demand correlation. As a result, a closed-form optimal solution is obtained which is an increasing function of degree of risk aversion.

Figure 2.4 also shows that consistent with my analysis in Sect. 2.7, negatively correlated demands result in higher optimal order quantities than independent demands under risk aversion, while positively correlated demand leads to lower

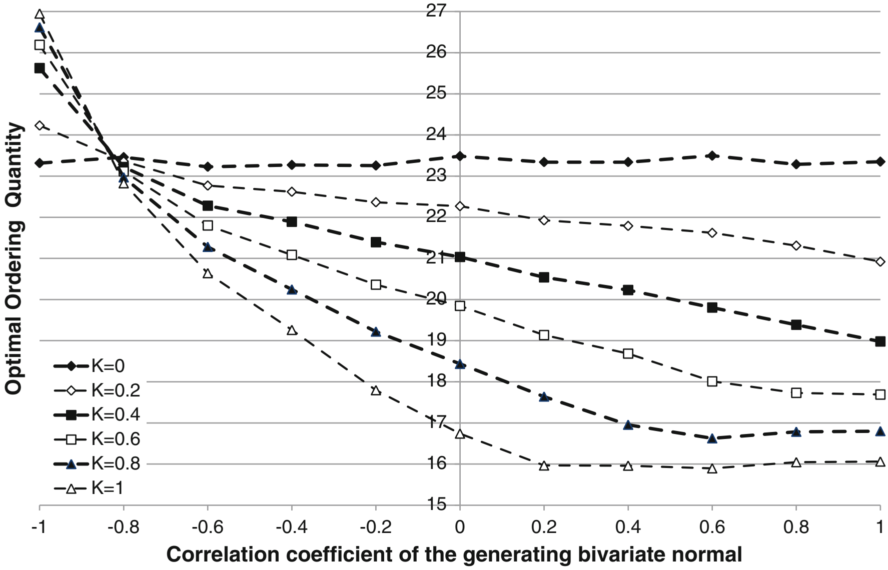


Fig. 2.4 Identical products with dependent demands—The impact of demand correlation and risk aversion κ

optimal order quantity under risk aversion. Indeed, the impact of demand correlation is almost monotonic with small deviations due to random sample errors.

These observations imply that if the firm is risk-averse, then demand dependence can have a significant impact on its optimal order quantities. They agree with the intuition that stronger positively (negatively) correlated demands indicate higher (lower) risk, and therefore lead to lower (higher) order quantities. More interestingly, while in most cases, the order quantity decreases in the degree of risk aversion, it can increase when the demands are strongly negatively correlated.

For heterogenous products, I consider a simple system with two products and the following parameters: $r_1 = 15, c_1 = 10, s_1 = 7$ and $r_2 = 30, c_2 = 10, s_2 = 4$. The demand is bivariate lognormal generated by exponentiating a bivariate normal with $\mu_1 = \mu_2 = 3$, $\sigma_1 = 0.4724$, $\sigma_2 = 1.26864$ and a correlation coefficient of $-1, -0.8, -0.6, \ldots, 1$. The marginal demand distributions of products 1 and 2 have means 22.46 and 44.913, standard deviations 11.23 and 89.826, and cv's 0.5 and 2, respectively. Intuitively, product 1 is less risky and less profitable than product 2.

My numerical study shows that for product 1, the impact of demand correlation is similar to that for identical products; see Fig. 2.5. For product 2, however, the optimal ordering quantity always decreases in κ but not in correlation, see Fig. 2.6.

The implication is that for heterogenous products, the impact of demand correlation under risk aversion can be very different in each product. Specifically, as the firm becomes more risk-averse, it should always order less of the more risky and more profitable products. However, for the less risky and less profitable products,

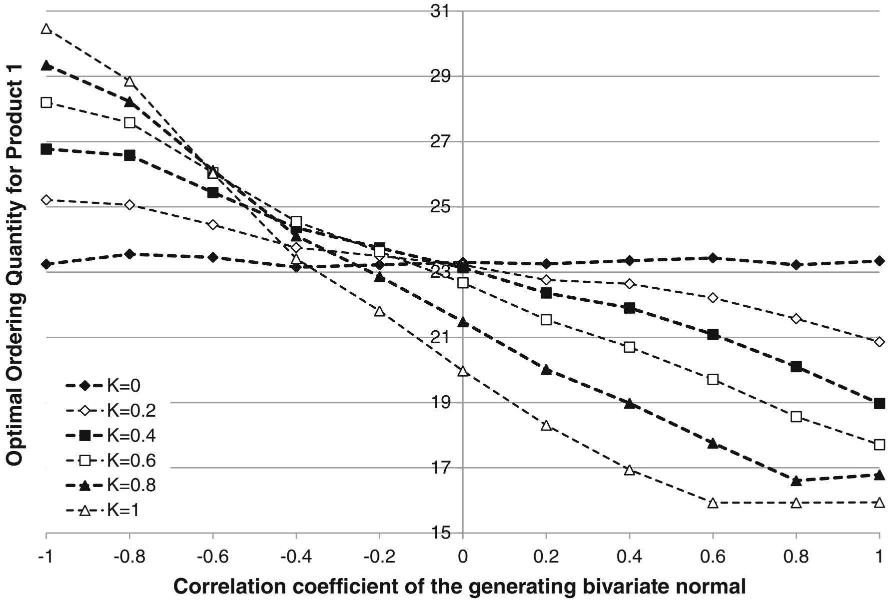


Fig. 2.5 Heterogenous products with dependent demands—The impact of demand correlation and risk aversion κ for the product with low risk and low profit

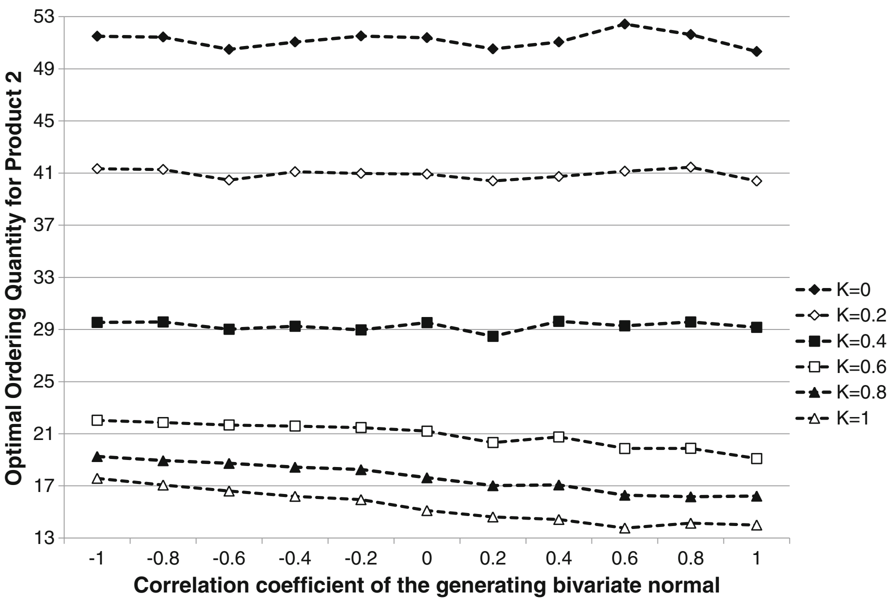


Fig. 2.6 Heterogenous products with dependent demands—The impact of demand correlation and risk aversion κ for the product with high risk and high profit

while it should order less when demands are positively correlated, it may order more when demands are strongly negatively correlated.

For more details on the numerical study, see Choi (2009).

2.9 Conclusions

The multi-product newsvendor problem with coherent measures of risk does not decompose into independent problems, one for each product. The portfolio of products has to be considered as a whole. My analytical results focus on the impact of risk aversion and demand dependence on the optimal order quantities. I analyze the asymptotic behavior of the optimal risk-averse solution. Then I derive (2.30) and (2.36) for general law-invariant coherent measures of risk, which are simple and accurate approximations of the optimal order quantities for a large number of products with independent demands. My numerical study confirms the accuracy of these approximations for the numbers of products as small as 10, and enriches my understanding of the interplay of demand dependence and risk aversion.

It is perhaps appropriate to conclude this paper by comparing the multi-product risk-averse newsvendor problem (2.4) to the risk-averse portfolio optimization problem. In a portfolio problem, one has n assets with random returns $R_1,\dots,R_n$ and the objective is to determine investment quantities $x_1,\dots,x_n$ to obtain desirable characteristics of the total portfolio return $P(x,R) = R_1x_1 + \cdots + R_nx_n$. In the classical mean-variance approach of Markowitz (1959), the mean of the return and its variance are used to find efficient portfolio allocations. See also Elton et al. (2006). In more modern approaches (e.g., Konno and Yamazaki 1991; Miller and Ruszczyński 2008; Ruszczyński and Vanderbei 2003) more general mean–risk models and coherent measures of risk are used, similarly to problem (2.4). There are, however, fundamental structural differences which make the multi-product newsvendor problem significantly different from the financial portfolio problem.

The most important difference is that the portfolio return $P(x,R)$ is *linear* with respect to the decision vector x, while the newsvendor profit $\Pi(x,D)$ is *concave* and *nonlinear* with respect to the order quantities x. This leads to the following different properties of the problems.

- The risk-neutral portfolio problem has no solution, unless the total amount invested (e.g., to 1) is restricted, in which case the optimal solution is to invest everything in the asset(s) having highest expected returns. On the contrary, the risk-neutral newsvendor problem always has a solution, because of natural limitations of the demand.
- The effect of using risk measures in the portfolio problem is a diversification of the solution, which otherwise would remain completely nondiversified. In the newsvendor problem the use of risk measures results in changes of the already diversified risk-neutral solution, by ordering more of products having less variable or negatively correlated demands and less of products having more

variable or positively correlated demands. Products are unlikely to be eliminated because of risk aversion, because very small amounts will almost always be sold and thus they introduce very little risk.

- In the portfolio problem, independently of the number of assets considered, the risk-neutral solution remains structurally different from the risk-averse solution. On the contrary, in the newsvendor problem the risk-neutral solution is asymptotically optimal under risk aversion, when the number of independent products approaches infinity.

Finally, it is worth stressing that the nonlinearity of the newsvendor profit $\Pi(x, D)$ is the source of formidable technical difficulties in the analysis of the composite function (2.4), which involves two nondifferentiable functions.

References

Acerbi, C., & Tasche, D. (2002). On the coherence of expected shortfall. *Journal Of Banking and Finance*, *26*(7), 1487–1503.

Ağrali, S., & Soylu, A. (2006). Single period stochastic problem with CVaR risk constraint. Presentation, Department of Industrial and Systems Engineering, University of Florida, Gainesville, FL.

Agrawal, V., & Seshadri, S. (2000a). Impact of uncertainty and risk aversion on price and order quantity in the newsvendor problem. *Manufacturing and Service Operations Management*, *2*(4), 410–423.

Agrawal, V., & Seshadri, S. (2000b). Risk intermediation in supply chains. *IIE Transactions*, *32*, 819–831.

Ahmed, S., Çakmak, U., & Shapiro, A. (2007). Coherent risk measures in inventory problems. *European Journal of Operational Research*, *182*, 226–238.

Ahmed, S., Filipović, D., & Svindland, G. (2008). A note on natural risk staistics, *Operations Research Letters*, *36*, 662–664.

Anvari, M. (1987). Optimality criteria and risk in inventory models: the case of the newsboy problem. *The Journal of the Operational Research Society*, *38*(7), 625–632.

Artzner, P., Delbaen, F., Eber, J., & Heath, D. (1999). Coherent measures of risk. *Mathematical Finance*, *9*(3), 203–228.

Aviv, Y., & Federgruen, A. (2001). Capacitated multi-item inventory systems with random and fluctuating demands: implications for postponement strategies. *Management Science*, *47*(4), 512–531.

Bonnans, J. F., & Shapiro, A. (2000). *Perturbation analysis of optimization problems*. New York, NY: Springer.

Bouakiz, M., & Sobel, M. (1992). Inventory control with an exponential utility criterion. *Operations Research*, *40*(3), 603–608.

Chen, F., & Federgruen, A. (2000). Mean-variance analysis of basic inventory models. Working paper, Division of Decision, Risk and Operations, Columbia University, New York, NY.

Chen, X., Sim, M., Simchi-Levi, D. & Sun, P. (2007). Risk aversion in inventory management. *Operations Research*, *55*, 828–842.

Chen, Y., Xu, M., & Zhang, Z. (2009). A risk-averse newsvendor model under the CVaR criterion. *Operations Research*, *57*(4), 1040–1044.

Cheridito, P., Delbaen, F., & Kupper, M. (2006). Dynamic monetary risk measures for bounded discrete-time processes. *Electronic Journal of Probability*, *11*, 57–106.

Choi, S. (2009). Risk-averse newsvendor models. PhD Dissertation, Department of Management Science and Information Systems, Rutgers University, Newark, NJ.

Choi, S., Ruszczyński, A., & Zhao, Y. (2011). A multiproduct risk-averse newsvendor with law-invariant coherent measures of risk. *Operations Research*, *59*(2), 346–364.

Chung, K. H. (1990). Risk in inventory models: the case of the newsboy problem – optimality conditions. *The Journal of the Operational Research Society*, *41*(2), 173–176.

Decroix, G., & Arreola-Risa, A. (1998). Optimal production and inventory policy for multiple products under resource constraints. *Management Science*, *44*, 950–961.

Delbaen, F. (2002). Coherent risk measures on general probability space. *Advance in finance and stochastics* (pp. 1–37). Berlin: Springer.

Dentcheva, D., & Ruszczyński, A. (2003). Optimization with stochastic dominance constraints. *SIAM Journal on Optimization*, *14*(2), 548–566.

Eeckhoudt, L., Gollier, C., & Schlesinger, H. (1995). The risk-averse (and prudent) newsboy. *Management Science*, *41*(5), 786–794.

Elton, E. J., Gruber, M. J., Brown, S. J., & Goetzmann, W. N. (2006). *Modern portfolio theory and investment analysis*. New York: Wiley.

Esary, J. D., Proschan, F., & Walkup, D. W. (1967). Association of random variables, with applications. *Annals of Mathematical Statistics*, *38*, 1466–1474.

Evans, R. (1967). Inventory control of a multiproduct system with a limited production resource. *Naval Research Logistics*, *14*(2), 173–184.

Federgruen, A., & Zipkin, P. (1984). Computational issues in an infinite-horizon, multi-echelon inventory model. *Operations Research*, *32*, 818–836.

Föllmer, H., & Schied, A. (2002). Convex measures of risk and rrading constraints. *Finance and Stochastics*, *6*, 429–447.

Föllmer, H., & Schied, A. (2004). *Stochastic finance: an introduction in discrete time*. Berlin: de Gruyter.

Gan, X., Sethi, S. P., & Yan, H. (2004). Coordination of supply chains with risk-averse agents. *Production and Operations Management*, *14*(1), 80–89.

Gan, X., Sethi, S. P., & Yan, H. (2005). Channel coordination with a risk-neutral supplier and a downside-risk-averse retailer. *Production and Operations Management*, *13*(2), 135–149.

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

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

Hadar, J., & Russell, W. (1969). Rules for ordering uncertain prospects. *The American Economic Review*, *59*, 25–34.

Hadley, G., & Whitin, T. M. (1963). *Analysis of inventory systems*. Englewood Cliffs: Prentice-Hall.

Heyde, C. C., Kou, S. G., & Peng, X. H. (2006). What is a good risk measure: bridging the gaps between data, coherent risk measures and insurance risk measures. Working paper, Department of Statistics, Columbia University, New York, NY.

Heyman, D. P., & Sobel, M. J. (1984). *Stochastic models in operations research: Volume 2. Stochastic optimization*. New York: McGraw Hill Company.

Howard, R. A. (1988). Decision analysis: practice and promise. *Management Science*, *34*, 679–695.

Ignall, E., & Veinott Jr., D. P. (1969). Optimality of myopic inventory policies for several substitute products. *Management Science*, *15*, 284–304.

Konno, H., & Yamazaki, H. (1991). Mean–absolute deviation portfolio optimization model and its application to Tokyo stock market. *Management Science*, *37*, 519–531.

Kusuoka, S. (2003). On law invariant coherent risk measures. *Advances in Mathematical Economics*, *3*, 83–95.

Kusuoka, S., & Morimoto, Y. (2004). Homogeneous law invariant coherent multiperiod value measures and their limits. Working Paper, Graduate School of Mathematical Science, University of Tokyo, Tokyo, Japan.

Lau, H. S. (1980). The newsboy problem under alternative optimization objectives. *The Journal of the Operational Research Society*, *31*(6), 525–535.

Lau, H. S., & Lau, H. L. (1999). Manufacturer's pricing strategy and return policy for a single-period commodity. *European Journal of Operational Research*, *116*, 291–304.

Lehmann, E. (1955). Ordered families of distributions. *Annals of Mathematical Statistics*, *26*, 399–419.

Levy, H. (2006). *Stochastic dominance: Investment decision making under uncertainty* (2nd ed.). New York: Springer.

Markowitz, H. M. (1959). *Portfolio Selection*. New York: Wiley.

Martinez-de-Albeniz, V., & Simchi-Levi, D. (2006). Mean-variance trade-offs in supply contracts. *Naval Research Logistics*, *53*, 603–616.

Miller, N., & Ruszczyński, A. (2008). Risk-adjusted probability measures in portfolio optimization with coherent measures of risk. *European Journal of Operational Research*, *191*, 193–206.

Müller, A., & Stoyan, D. (2002). *Comparison methods for stochastic models and risks*. Chichester: Wiley.

Ogryczak, W., & Ruszczyński, A. (1999). From stochastic dominance to mean-risk models. *European Journal of Operational Research*, *116*, 33–50.

Ogryczak, W., & Ruszczyński, A. (2001). On consistency of stochastic dominance and mean-semideviation models. *Mathematical Programming*, *89*, 217–232.

Ogryczak, W., & Ruszczyński, A. (2002). Dual stochastic dominance and related mean-risk models. *SIAM Journal on Optimization*, *13*(1), 60–78.

Porteus, E. L. (1990). Stochastic inventory theory. In D. P. Heyman & M. J. Sobel (Eds.), *Handbook in OR and MS* (Vol. 2, Chapter 12, pp. 605–652). North-Holland: Elsevier Science Publishers B.V.

Pospišil, L., Večer, J., & Xu, M. (2008). Tradeable measures of risk. Working Paper, Department of Statistics, Columbia University, New York, NY.

Prékopa, A. (2003). Probabilistic programming. In A. Ruszczyński & A. Shapiro (Eds.), *Stochastic programming*. Amsterdam: Elsevier.

Riedel, F. (2004). Dynamic coherent risk measures. *Stochastic Processes and their Applications*, *112*, 185–200.

Rockafellar, R. T., & Uryasev, S. P. (2000). Optimization of conditional value-at-risk. *The Journal of Risk*, *2*, 21–41.

Ruszczyński, A., & Shapiro, A. (2005). Optimization of risk measures. In G. Calafiore & F. Dabbene (Eds.), *Probabilistic and randomized methods for design under uncertainty* (pp. 117–158). London: Springer.

Ruszczyński, A., & Shapiro, A. (2006a). Optimization of convex risk functions. *Mathematics of Operations Research*, *31*(3), 433–452.

Ruszczyński, A., & Shapiro, A. (2006b). Conditional risk mappings. *Mathematics of Operations Research*, *31*(3), 544–561.

Ruszczyński, A., & Vanderbei, R. J. (2003). Frontiers of stochastically nondominated portfolios. *Econometrica*, *71*(4), 1287–1297.

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

Shapiro, A. (2008). Asymptotics of minimax stochastic programs. *Statistics and Probability Letters*, *78*, 150–157.

Shapiro, A., Dentcheva, D., & Ruszczyński, A. (2009). *Lectures on stochastic programming: Modeling and theory*. Philadelhpia: SIAM (Society for Industrial and Applied Mathematics).

Tomlin, B., & Wang, Y. (2005). On the value of mix flexibility and dual sourcing in unreliable newsvendor networks. *Manufacturing and Service Operations Management*, *7*(1), 37–57.

Tong, Y. L. (1980). *Probability inequalities in multivariate distributions*. New York: Academic.

Tsay, A. A. (2002). Risk sensitivity in distribution channel partnerships: implications for manufacturer return policies. *Journal of Retailing*, *78*, 147–160.

van Mieghem, J. A. (2003). Capacity management, investment and hedging: review and recent developments. *Manufacturing Service Operations Management*, *5*(4), 269–302.

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

van Ryzin, G. J., & Mahajan, S. (1999). On the relationship between inventory cost and variety benefits in retail assortments. *Management Science*, *45*, 1496–1509.

Veinott, A. (1965). Optimal policy for a multi-product, dynamic, nonstationary inventory problem. *Management Science*, *12*, 206–222.

von Neumann, J., & Morgenstern, O. (1944). *Theory of games and economic behavior*. Princeton: Princeton University Press.

Wang, S. S., Young, V. R., & Panjer, H. H. (1997). Axiomatic characterization of insurance prices. *Insurance: Mathematics and economics*, *21*, 173–183.

Chapter 3
A Copula Approach to Inventory Pooling Problems with Newsvendor Products

Burcu Aydın, Kemal Guler, and Enis Kayış

Abstract This study focuses on the inventory pooling problem under the newsvendor framework. The specific focus is the change in inventory levels when product inventories are pooled. We provide analytical conditions under which an increase (or decrease) in the total inventory levels should be expected. We introduce the copula framework to model a wide range of dependence structures between pooled demands, and provide a numerical study that gives valuable insights into the effects of marginal demand distributions and dependence structure on inventory pooling decisions.

Keywords Copula approach • Inventory pooling • Multiple sources • Total inventory levels • Sklar's Theorem

3.1 Introduction: Inventory Pooling Problem

We study the inventory pooling problem using the classic newsvendor framework. The newsvendor problem occurs when for a given item, the inventory level is decided before the realization of the demand. Therefore, the optimal inventory level needs to be decided based on the distribution of the stochastic demand *D*. Unsold items at the end of the period are typically assumed to be either discarded or salvaged. The solution of the newsvendor problem is well known: a quantile of the demand distribution depending on price and cost of the item is the stock level that is optimal in terms of profit.

The pooling problem occurs when the decision makers have the option to combine inventories for an item that serves multiple demand sources. The pooling could be in the form of determining one physical inventory holding location

B. Aydın • K. Guler • E. Kayış (✉)
HP Labs, 1501 Page Mill Rd Palo Alto, CA 94304, USA
e-mail: aydin@hp.com; kemal.guler@hp.com; enis.kayis@hp.com

T.-M. Choi (ed.), *Handbook of Newsvendor Problems*, International Series in Operations Research & Management Science 176, DOI 10.1007/978-1-4614-3600-3_3,

that will serve multiple locations, setting up quick transshipment modes between different inventory locations (therefore allowing to plan inventories together), or even designing two products so that they are substitutable if the need arises. This effort has a clear reward: It is a well-known fact in the literature that pooling always leads to higher profits. (see, for example, Corbett and Rajaram 2006). However, the optimal inventory levels in the system may increase or decrease after pooling.

Our paper primarily focuses on the change in optimal inventory levels when demands from multiple sources are pooled. The change in inventory levels is an important decision factor. Contrary to common intuition, pooling may result in a decrease or increase in total inventory levels. The pooling decision may bring additional costs that depend on the targeted inventory level. The costs could be due to adjusting warehouse capacities, redesign costs, etc., and profits may include reduced stock-out rates and therefore higher customer satisfaction. These should be carefully weighed together with the profit increase due to pooling. Furthermore, after the pooling decision is made, adjusting the inventory levels to the new optimal levels is important in achieving higher profits. The new optimal levels depend on the demand for the product in each channel, and how these channels affect each other. We investigate how pooled inventory levels are affected by marginal distributions of product demands and the dependence structure between them.

The question we tackle in this paper is mentioned, though not solved, by Corbett and Rajaram (2006):

> Most of this literature in inventory pooling, ... , focuses on the impact of pooling on expected profits. A related, but usually more intractable problem, concerns the effect of pooling on optimal inventory levels. We do not consider that question here, though some work, including Eppen (1979), Erkip et al. (1990), and Van Mieghem and Rudi (2002) do address that issue under more restrictive distributional assumptions than ours. So far, the work related to pooling of inventories has generally lacked a formal mechanism for assessing the impact of dependence on the value of pooling when demands are nonnormal. Whenever dependence has been explicitly included, it has generally been in the context of bivariate or multivariate normal demands.

Other previous studies on the subject handled certain cases where the demand distributions of the channels and their relationship can be explained by well-known multi-variate distributions. Often independent and identically distributed (IID) demand is assumed. For example, Gerchak and Mossman (1992) assume IID exponentially distributed marginals, and Yang and Schrage (2009) studies IID right-skewed marginals. This approach provides mathematical tractability. In real-life applications, however, product demands are neither identical nor independent. In this paper, we take up this problem and show that the theory of copulas provides a powerful and tractable yet rigorous framework to address the effect of relaxing both independent and identical demand assumptions on the optimal pooled inventory levels. They also allow us to analyze a very wide range of different dependence structures that may not fit into any of the well-known multi-variate distributions.

The details of the model we consider are as follows. The cost of stocking each unit is c. For each demand unit that can be satisfied from inventory, a revenue of p is made. Unsatisfied demand is lost as well as the overstocked items. The objective is to decide the inventory level Q that will maximize the expected total profit. It is well known that the optimal inventory level is a quantile of the demand distribution, i.e.:

$$F^{-1}\left(\frac{p-c}{p}\right) = arg\max_{Q}\left\{pE_D\left[\min(D,Q)\right] - cQ\right\}, \tag{3.1}$$

where $F(.)$ is the distribution function of demand.

In the stylized inventory pooling problem, two identical items with uncertain demands, D_1 and D_2, are considered. These items have the same unit profit and unit stocking cost. The decision maker has two options: Keeping a dedicated inventory to satisfy the demand of each item, or holding a single inventory for the aggregate demand, $D_1 + D_2$. It has been shown that pooling is a better option; however, one still needs to decide on the optimal inventory levels. In the first option, the optimal inventory in the system can be shown to be $F_1^{-1}(t) + F_2^{-1}(t)$, where $F_i(.)$ is the marginal distribution function of D_i and $t := \frac{p-c}{p}$ is defined as the margin ratio. It is easy to see that this quantity is independent of the dependence structure between the demands. On the other hand, the optimal pooled inventory level, $F_{1+2}^{-1}(t)$, depends not only on the marginal demand distributions but also on the dependence structure between D_1 and D_2 (where $F_{1+2}(x) := Pr(D_1 + D_2 \leq x)$).

From a practical point of view, the manager knows pooling is a better option, but he needs to decide whether to keep more or less total inventory as a result of that decision. If pooling requires higher levels of total inventory, we say that *pooling effect is positive*. Similarly, *pooling effect is negative* when pooled inventory level is lower than the dedicated inventory. In other words, we define the pooling effect as $F_{1+2}^{-1}(t) - F_1^{-1}(t) - F_2^{-1}(t)$.

3.2 Literature Review

The inventory pooling problem has been studied extensively in the operations management literature. For many of these studies, the main focus has been the profit comparison under various settings. A smaller number of studies take up the problem of determining the pooled inventory levels.

The earliest and most well-known reference on the pooling problem is Eppen (1979). This study considers the pooling problem when product demands are jointly distributed with multi-variate normal distribution with a known covariance matrix. He shows that the centralized system brings cost savings, and the magnitude of these savings depend on the correlation: the lower the dependence, the higher the savings.

Since Eppen (1979), costs and benefits of inventory pooling are investigated under various other settings. See Gerchak and He (2003) and Alfaro and Corbett (2003) for recent reviews.

Netessine and Rudi (2003) focus on the inventory centralization problem for substitutable products. Substitution is the technical equivalent of pooling when full substitution without stock out penalties are allowed. They show that, when substitution is allowed, it is possible that the optimal inventory levels may increase for some items that are being pooled. However, they give results on the levels of individual items, and they do not provide any result on the total inventory level of the items being pooled under centralization.

Erkip et al. (1990) takes up a similar question: the centralization of inventory ordering policies under the newsvendor framework. They investigate the effect of correlation between normally distributed demands of items that can be centralized. They conclude that "the effect of correlation can be highly significant, resulting in significantly larger amounts of safety stock for optimal control compared to the no-correlation case."

In their 2003 paper, Gerchak and He investigate the effect of demand variances on the pooling savings. They provide a framework in which an increase in demand variability always increases the savings achieved by combining these demands. They do not require the demand distributions to be independent for their result to hold. However, they do not study how the combined inventory levels are affected by variability.

Alfaro and Corbett (2003) ask an interesting question: if the inventory levels are not optimal in a current setting, would pooling still bring savings? They investigate the profits coming from pooling under nonoptimal inventory levels, and compare this to the benefit of optimizing the separate inventory levels rather than pooling them. They find conditions under which it is better to optimize the inventory levels of dedicated setting, and conditions where pooling only will be more profitable.

In inventory pooling literature, the effect of dependence on the optimal inventory levels has been studied assuming multi-variate normal demands. Corbett and Rajaram (2006) use copulas to model the dependence structure between demands. As noted by the authors, they focus on the impact of pooling on expected profits. This focus is particularly essential as the results of superiority of pooling rely critically on the ability to find the optimal inventory levels.

The small number of studies that focus on the pooled inventory levels provide examples in which pooling leads to higher inventory levels contrary to the earlier intuition. For example, Pasternack and Drezner (1991) show that this comparison depends on the transfer revenue. Transfer revenue is the profit that comes from the substitution of one product when the other's inventory is depleted. Their cost structure for the substituted amount is different than the original costs; therefore, their results are not directly comparable to the studies where pooling is understood as full substitution, where costs do not change if parts are substituted.

Gerchak and Mossman (1992) conclude that, contrary to the prevalent intuition, pooling may lead to higher inventory levels when demands are IID with an exponential distribution and price per unit and the ratio of cost of underage to cost of overage is sufficiently low. They show this using a numerical counterexample where the demands have exponential distribution. However, they do not provide any generalized findings in terms of providing analytical conditions or distributions which would imply higher pooled inventory levels.

In a recent paper, Yang and Schrage (2009) define the case in which pooling increases inventory levels as “inventory anomaly.” They focus on IID right-skewed demand distributions as marginals. They claim that for any two IID right-skewed demand distributions, there exists a range of the margin ratio $\frac{p-c}{p}$ where pooling leads to higher inventory levels. Moreover, for any newsvendor ratio $\frac{p-c}{p} \geq 0.5$, a right-skewed distribution of IID marginals that leads to higher pooled inventory levels exist. Their result is important in the sense that they describe certain conditions where the “inventory anomaly” can be expected.

Our paper attempts to provide a much more general framework, where the level of pooled inventory can be found under any demand distribution and dependence structure through the use of copulas. Our numerical analysis provides examples of some well-known copulas and marginal distributions that can be used.

3.3 Comparison of Inventory Levels

In this section, we explore a qualitative question: How does the sign of pooling effect change as the margin ratio t varies? The following Proposition sheds light into this question:

Proposition 1. *Let* $P(t) = F_{1+2}^{-1}(t) - F_1^{-1}(t) - F_2^{-1}(t)$. *Assume that* $P(t)$ *has a unique root* t_0 *in* $(0,1)$. *Then,* t_0 *is a threshold such that the pooling effect is negative at* t *if and only if* $t > t_0$.

Proof of Proposition 1 follows from Theorem 1 of Liu and David (1989). The proposition stipulates that if $P(t)$ has a unique root in $(0,1)$, then pooling effect can only change sign from negative to positive as t increases, the critical threshold being t_0. This threshold depends on both the marginals and the dependence structure of joint demand distribution. However, we can characterize this value under certain settings. First, it is easy to verify that for multi-normal family of demand distributions, this critical threshold is always 0.5. That is, regardless of specific parameters that describe a multi-normal demand, pooling leads to a lower inventory level if and only if margin ratio is higher than 0.5. One can extend this “detail-free” threshold result to other distributions from the same family.

Proposition 2. *If* (D_1, D_2) *follows a distribution in the elliptical family,*[1] *then pooling leads to lower inventory if and only if the margin ratio is higher than* 0.5.

The proof of this proposition follows trivially from Theorem 6.8 of McNeil et al. (2005).

[1]The elliptical family includes well-known distributions such as Normal, Laplace, Student-t, Cauchy, and Logistic among others.

For the next result, we use the following definition:

Definition 1. A distribution function F is regularly varying at minus infinity with tail index $\alpha > 0$ if,

$$\lim_{t\to\infty} \frac{F(-tx)}{F(-t)} = x^{-\alpha} \quad \forall x > 0.$$

The next proposition shows that if the distribution of demand is regularly varying, then the sign of pooling effect depends on the tail properties of the demand distribution.

Proposition 3. *Assume that the tail probability of the joint demand distribution to be negligible compared to those of marginal demand distributions. Moreover, let D_1 and D_2 are identically distributed with regularly varying distribution functions with the same tail index α. There exists a threshold $0 < t_0 < 1$ such that if the margin ratio t is greater than or equal to t_0 then:*

- *The pooling effect is negative if $\alpha > 1$.*
- *The pooling effect is positive if $0 < \alpha < 1$.*

The proof is from Theorem 10 of Jang and Jho (2007).

It is possible to have $t_0 = 0$ which implies that pooling leads to higher inventory for all margin ratios. For example, when demands are IID with Pareto distributions, which has infinite mean, then the threshold becomes 0.

Having established some conditions for positive and negative pooling effect and that pooling may lead to either higher or lower inventory levels, we next investigate the effect of characteristics of the demand uncertainty on inventory levels. With multiple products, we need to study the effect of the marginal demand distributions as well as the dependence structure between these demands. Toward this end, copula representation provides a unified and rigorous approach for which we provide a short overview.

3.4 Brief Overview of Copula Theory

The joint distribution of demand is critical in understanding the behavior of optimal pooled inventory levels. There are two components of a joint distribution: the marginal distributions of each demand source, and the dependence between these demand sources. In order to study the effect of these components independently, we introduce the copula theory.[2] Copulas join the univariate marginal distributions of individual random variables to arrive at the joint distribution function for these variables.

[2]Nelsen (1999) provides an excellent general introduction to the theory of copulas.

Definition 2. A d-dimensional copula $C(u_1,\ldots,u_d)$ is a distribution function on $[0,1]^d$ with standard uniform marginal distributions.

McNeil et al. (2005) shows that a function C with the following properties is a copula:

1. $C(u_1,\ldots,u_d)$ is increasing in each component u_i.
2. $C(1,\ldots,1,u_i,1,\ldots,1) = u_i$ for all $i \in \{1,\ldots,d\}$, $u_i \in [0,1]$.
3. For all $(a_1,\ldots,a_d)$, $(b_1,\ldots,b_d)$ in $[0,1]^d$ with $a_i \leq b_i$, we have:

$$\sum_{i_1=1}^{2} \cdots \sum_{i_d=1}^{2} (-1)^{i_1+\cdots+i_d} C(u_{1i_1},\ldots,u_{di_d}) \geq 0,$$

where $u_{j1} = a_j$ and $u_{j2} = b_j$ for all $j \in \{1,\ldots,d\}$.

Sklar's Theorem shows that when the marginal distributions are continuous, then the copula is unique.

Theorem 1 (Sklar's Theorem). *Let $F(x_1,x_2,...,x_n)$ be an n-dimensional joint distribution with continuous marginals $F_1(x_1),...,F_n(x_n)$. Then the joint distribution has a unique copula representation given by*

$$F(x_1,x_2,...,x_n) = C(F_1(x_1),...,F_n(x_n)). \tag{3.2}$$

Sklar's Theorem provides a powerful technique that enables the separation of marginal distributions from the dependence structure. Since one can fix or vary the marginals and the copula separately, a rich class of stochastic models can be constructed. Copula-marginal representation of the joint distribution of a set of random variables has been used in a variety of application areas from decision and risk analysis to finance.

Some of the most commonly used examples of the copula include the product, Gaussian, and Archimedean family copulas. The product copula models the independent marginals case. The most commonly used multi-variate distribution, Normal, can be uniquely represented by normal marginals and a Gaussian copula.

Three important copulas within the Archimedean family are Gumbel, Clayton, and Frank. An important property that can be modeled using some Archimedean family copulas is the asymmetry around the mode. With these copulas, the dependence structure varies along the different section of the distribution tails. Two copulas that have this property are Gumbel and Clayton. Gumbel distribution is given by:

$$C_{\text{Gu}}(u_1,...,u_n) = e^{-((-ln(u_1))^{\theta}+...+(-ln(u_n))^{\theta})^{1/\theta}},$$

and Clayton copula is given by:

$$C_{\text{C}}(u_1,...,u_n) = \left(1-n+\sum_{i=1}^{n} u_i^{-1/\theta}\right)^{-\theta}.$$

Gumbel copula could be used when the dependence is higher in the right tail, and Clayton could be used when the dependence is higher in the left tail. Clayton copula exhibits higher dependence in the left tail. Finally, the Frank copula is given as:

$$C_{\mathrm{F}}(u_1,\ldots,u_n) = \log_\alpha \left(\frac{\prod_{i=1}^{n} \alpha^{u_i} - 1}{(\alpha - 1)^{(n-1)}} + 1 \right).$$

Frank copula is symmetric; it exhibits dependence on both tails. In our numerical analysis, we will focus on these three Archimedean copulas. The most important shape qualities of the popular Gaussian copula is already carried by the Frank copula. Moreover, Frank copula can represent negative dependence structures as well.

While all of these copula functions represent different two-dimensional structures, they can be compared through a summary scalar of dependence. One such scalar used commonly is called Kendall's τ. For a joint distribution, Kendall's τ is independent of the marginals and only depends on the copula. It varies between $[-1,1]$, while -1 represents perfect negative correlation, 0 represents lack of correlation and 1 represents perfect positive correlation. Given a two-dimensional copula function, the associated Kendall's τ can be found through the following formula:

$$\tau = 4 \int_0^1 \int_0^1 C(u,v)\mathrm{d}C(u,v) - 1. \tag{3.3}$$

This formula can be found in Kaas et al. (2009).

Other scalar measures of dependence also exist. Out of those, we do not use Pearson's r, since it only measures linear dependence and is not a robust measure of nonlinear dependence cases. Spearman's ρ and Blomqvist's β are two others that are also common and can measure nonlinear dependence. These could have been used instead of Kendall's τ. However, our numerical results would have come out very similar; therefore, we limited our analysis to Kendall's τ only. A wide discussion of these measures can be found in McNeil et al. (2005) and Nelsen (1999).

3.5 Numerical Analysis

In this section, we independently study the effect of identical versus nonidentical and symmetric versus asymmetric marginal demand uncertainties, as well as different types of copulas with varying forms and levels of tail dependence. Our focus is to investigate whether a unique threshold exists beyond which pooling leads to higher inventory levels, the value of this threshold, and the magnitude of the pooling effect. We connect our observations to managerial insights and complement the existing work on optimal pooled inventory levels.

The first factor in determining the inventory levels is the marginal distribution of each demand source. To understand the effect of skewness in marginal demand sources, we use the beta family. The support for the standard beta family is $[0,1]$; hence, the optimal total inventory level is always between 0 and 2. To study the effect of left or right skewness in the marginal demand, we use $\beta(2,8)$, and $\beta(8,2)$, respectively; we keep the variance fixed by only exchanging the two parameters of the distribution. The case with equal parameters ($\beta(5,5)$) represents the distribution example where the density is symmetric at both tails. Many sources use Normal distribution for this purpose which cannot model skewed cases. Another setting we will investigate is when the marginals are not identical.

The knowledge of marginal distributions is sufficient to determine the optimal dedicated inventory level. However, the optimal pooled inventory level also depends on the dependence structure. To understand the effect of dependence structure independent of the level of dependence, we fixed Kendall's τ to four different values (0, 0.2, 0.5, and 0.8) and computed the corresponding copula parameters using (3.3) for the copula families used. For the Frank copula that allows negative correlation, we followed the same procedure on the negative side as well as the positive side.

Figure 3.1 depicts the main copula functions that we will be using in our analysis: Gumbel, Clayton, and Frank. All of the these can model strong dependence as well as weak dependence.

Comparing the graphs of different copulas under the same Kendall's τ in Fig. 3.1, we can see that similar levels of "correlatedness" can exist in very different dependence structures. As Kendall's τ increases, the densities tend to concentrate around the 45° line. Gumbel copula is appropriate to model cases in which it is slightly more likely that high-level demands are correlated (i.e., higher the dependence on the right top quadrant). Clayton copula models cases where low-level demands are more correlated, perhaps due to unfavorable market conditions that affect all demand sources (i.e., higher the dependence on the left bottom quadrant). Frank copula, on the other hand, shows a more dispersed structure and models cases where dependence is similar in high and low level demands (i.e., it is symmetrical at both tails).

The combined affect of marginals and copulas is what drives the magnitude and sign of the pooling effect at any margin ratio. We will not give the joint density plots of all the marginal-copula pairs that will be used in the numerical analysis, but for illustrative purposes, the density plots belonging to Gumbel copula are given in Fig. 3.2. Other density plots reveal similar observations, so they will be omitted to save space.

To illustrate how both the pooled inventory level and the sum of dedicated inventory levels change under margin ratio, we present Fig. 3.3. This graph gives the intuition on how the dedicated and also pooled inventory levels change when margin ratio changes; i.e., when risk taking is more or less costly. We see that while the total dedicated inventory level steadily rises with respect to the margin ratio, pooled inventory level is more robust when margin ratio is medium and more sensitive

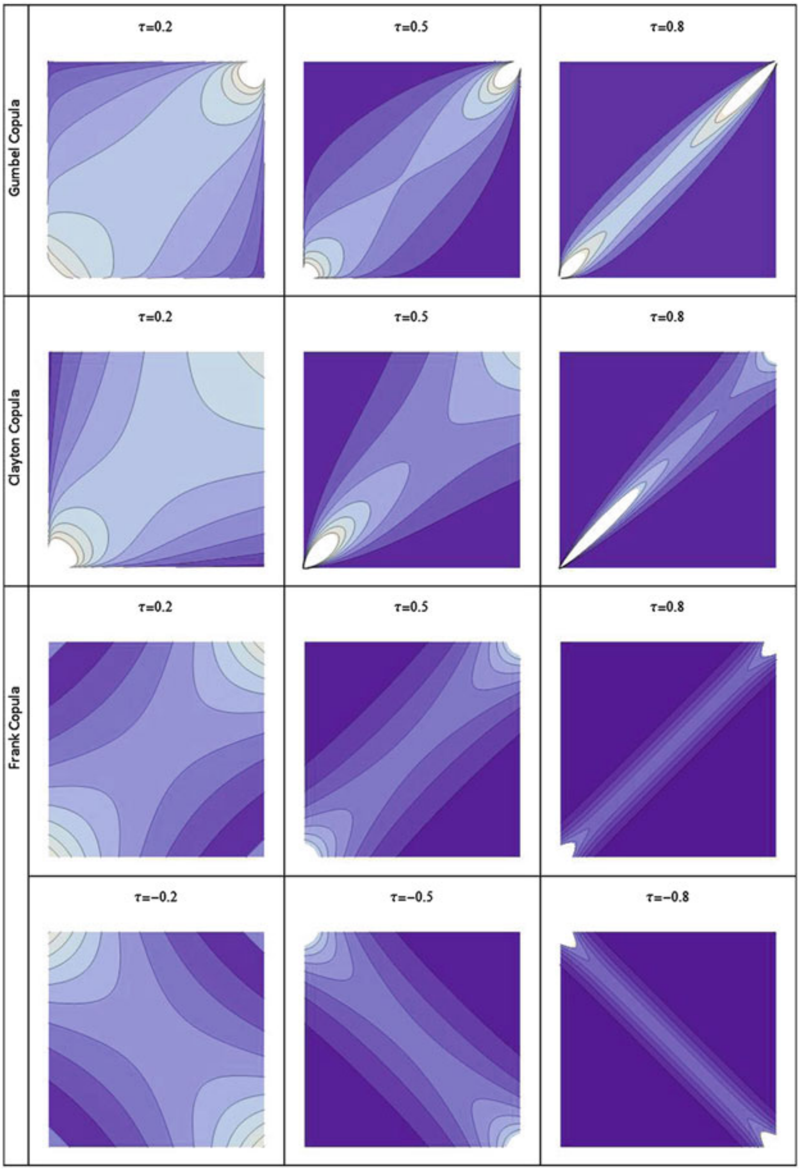


Fig. 3.1 Example dependence structures between two dependent random variables represented as contour plots of their copulas. Three Archimedean copula functions, Gumbel, Clayton, and Frank are shown. *Lighter shades* represent the areas with higher density, and *darker shades* represent areas with lower density. The function parameters are selected such that the copulas depict the dependence structures under Kendall's $\tau = 0.2$, 0.5, and 0.8 for positive dependence cases (first three rows), and Kendall's $\tau = -0.2$, -0.5, and -0.8 for negative dependence with Frank copula (fourth row)

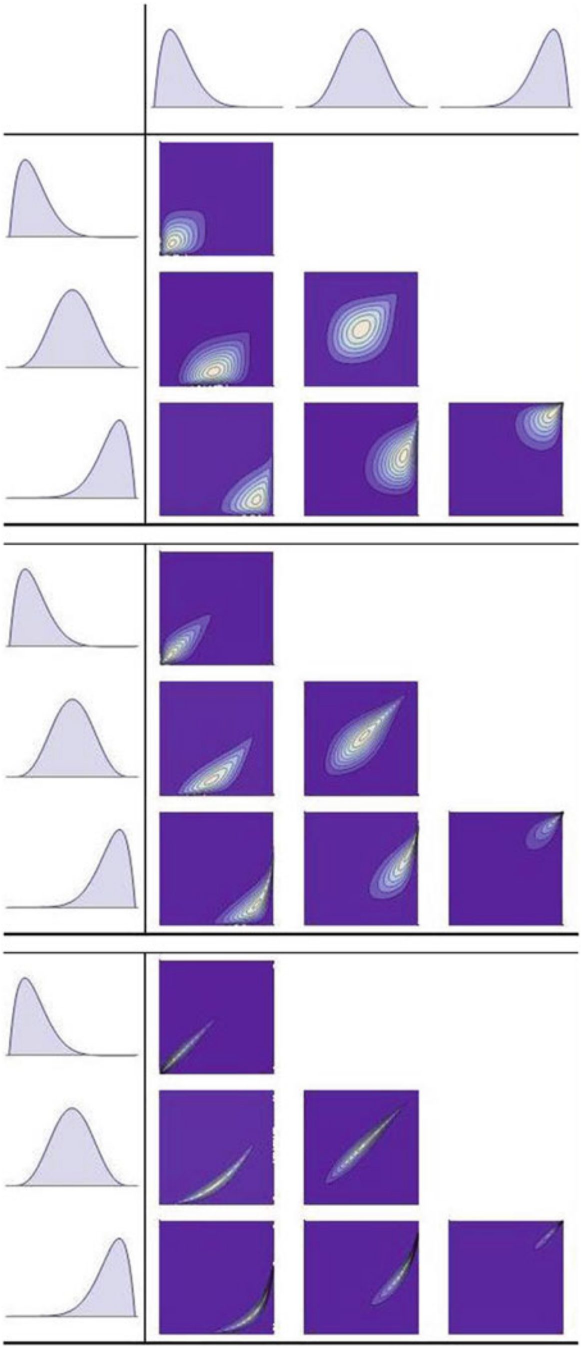


Fig. 3.2 The density functions of joint distributions obtained by combining Gumbel copula at three different Kendall's τ levels, and three different marginals ($\beta(2,8)$, $\beta(5,5)$, and $\beta(8,2)$). On *top row* and the *leftmost column*, the plots of marginal densities used are given. The *top panel* shows six different combinations of these marginals under the Gumbel copula with Kendall's $\tau = 0.2$. *Middle panel* shows the same combinations under Gumbel copula with $\tau = 0.5$, and *bottom panel* is $\tau = 0.8$. The effect of higher dependence can be seen from *top to bottom*: higher dependence concentrates the mass around its center, and reduces dispersion. The effect of nonsymmetry of Gumbel copula also becomes visible

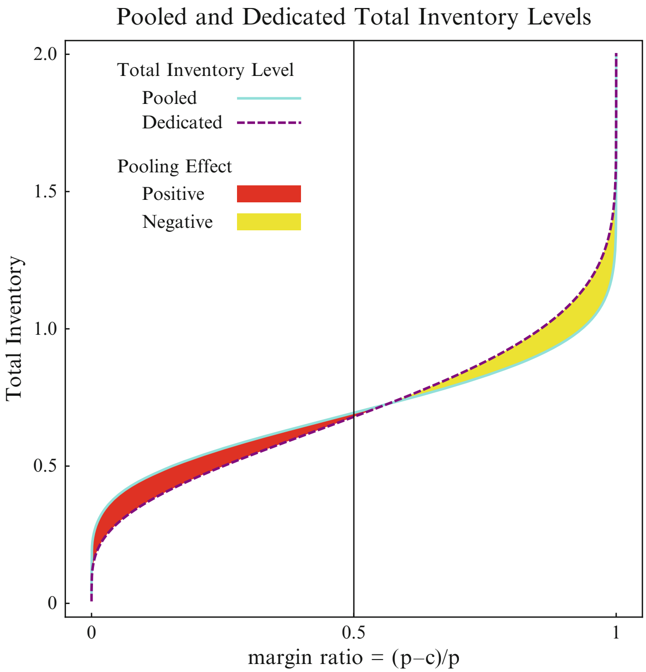


Fig. 3.3 The graph of dedicated (*dashed purple line*) and pooled inventory levels (*solid blue line*) where the horizontal axis represents the margin ratio. Product demands are identically distributed with Beta where $\alpha = 2$ and $\beta = 4$. Their dependence structure is given by Frank copula with parameter $\alpha = 100$ (Kendall's $\tau = -0.43$ for this parameter). The horizontal axis changes from 0 to 1 where 0.5 threshold is marked with a *vertical gray line*

when margin ratio is either too small or too high. The effects seen on this graph are consistent with the behaviors we observe in Figs. 3.4–3.7. Similar graphs under different copulas and marginals contain similar trends with respect to the shape of inventory level curves, so they will not be presented here.

3.5.1 Effect of Marginal Demand Distribution

The skewness of the marginal demand distribution affects both the threshold and the magnitude of pooling effect. First, the threshold t_0 may be either less or greater than the critical value 0.5 as the skewness changes. In general, left skewness in any marginal tends to increase the threshold, while right skewness has the opposite effect. This means that when a demand source is more likely to be low than high, then pooling leads to higher inventory for even smaller levels of the margin ratio. This follows from the fact that left skewed distributions concentrate the mass to lower left area of the joint distribution density, while right skewed ones concentrate on the upper right area.

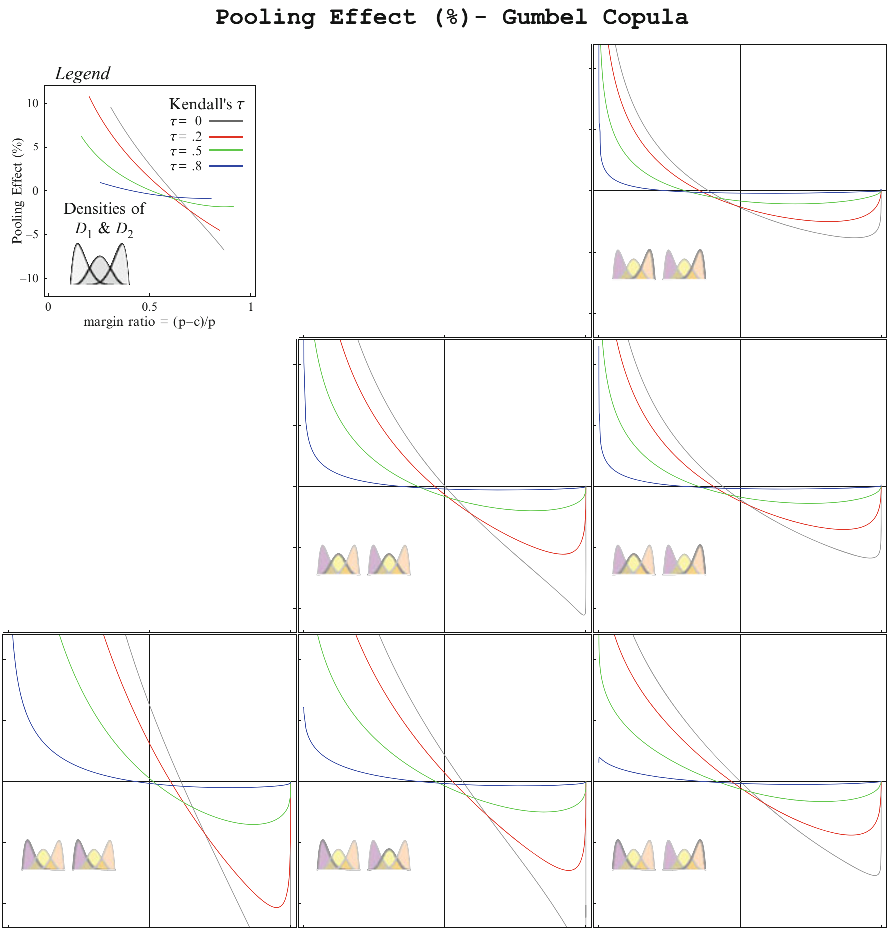


Fig. 3.4 The plots of percentage pooling effect on inventory levels under Gumbel copula. The marginal distributions are, from *top down*, $\beta(2,8)$, $\beta(5,5)$, and $\beta(8,2)$, and from *left to right*, $\beta(2,8)$, $\beta(5,5)$, and $\beta(8,2)$. These represent left-skewed, symmetrical, and right-skewed marginals. The common legend for subplots is given on the *top-left corner*

For two identical right-skewed marginals, it is clear from the plots that pooling effect is positive for small values of t. This is consistent with the result of Yang and Schrage (2009) on the existence of positive pooling effect for two IID right skewed distributions with small t. Our numerical results extend this finding to nonidentical marginal demands. We find that as one of the marginals changes from left skewed to right skewed while keeping the other one fixed, the threshold decreases, implying a positive pooling effect over a larger set of margin ratio. We also observe that the left skewness of marginals increases the threshold. Finally, we find that the case with one demand marginal being left-skewed and the other being right-skewed leads to similar results to the case wherein both marginals are symmetric around the mean.

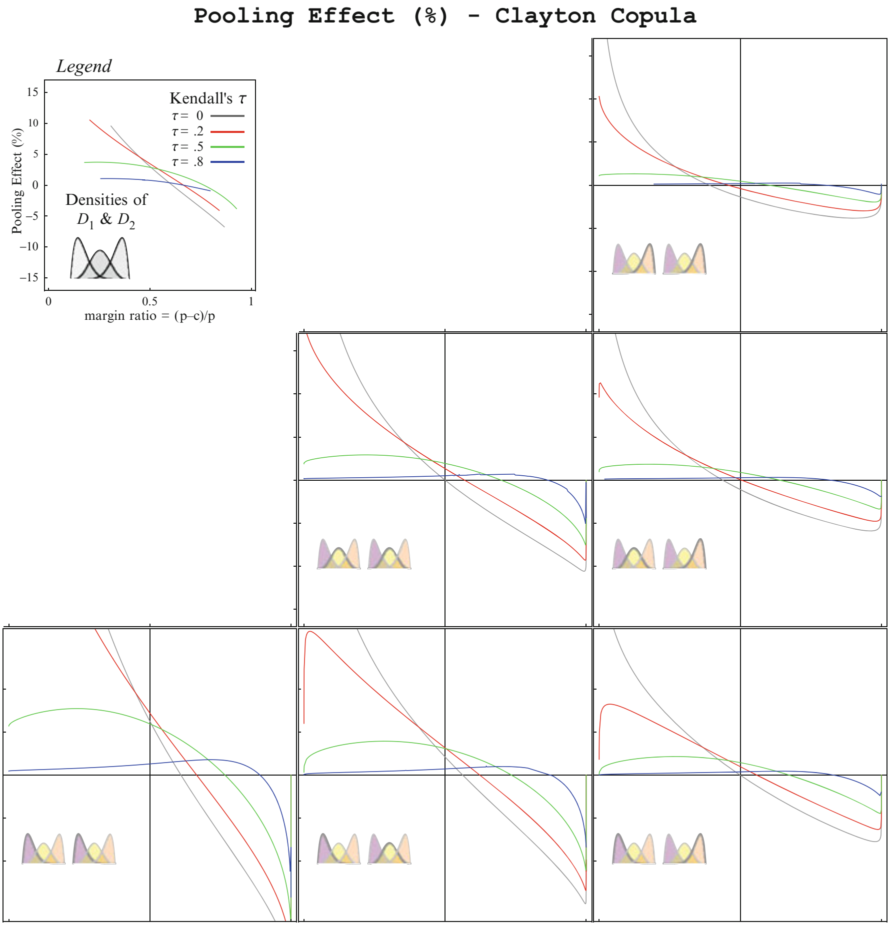


Fig. 3.5 The plots of percentage pooling effect on inventory levels under Clayton copula. The marginal distributions are, from *top down*, $\beta(2,8)$, $\beta(5,5)$, and $\beta(8,2)$, and from *left to right*, $\beta(2,8)$, $\beta(5,5)$, and $\beta(8,2)$. These represent left-skewed, symmetrical, and right-skewed marginals. The common legend for subplots is given on the *top-left corner*

The magnitude of the pooling effect decreases as marginals change from being left skewed to right skewed, for any or both of the marginals. When both of the marginals are left skewed we observe that pooling changes the inventory levels most, especially when margin ratio is small.

3.5.2 *Effect of Dependence Structure*

A comparison across the different copulas gives us interesting insights into seeing the effect of dependence structure, apart from the level of dependence itself, on the inventory levels.

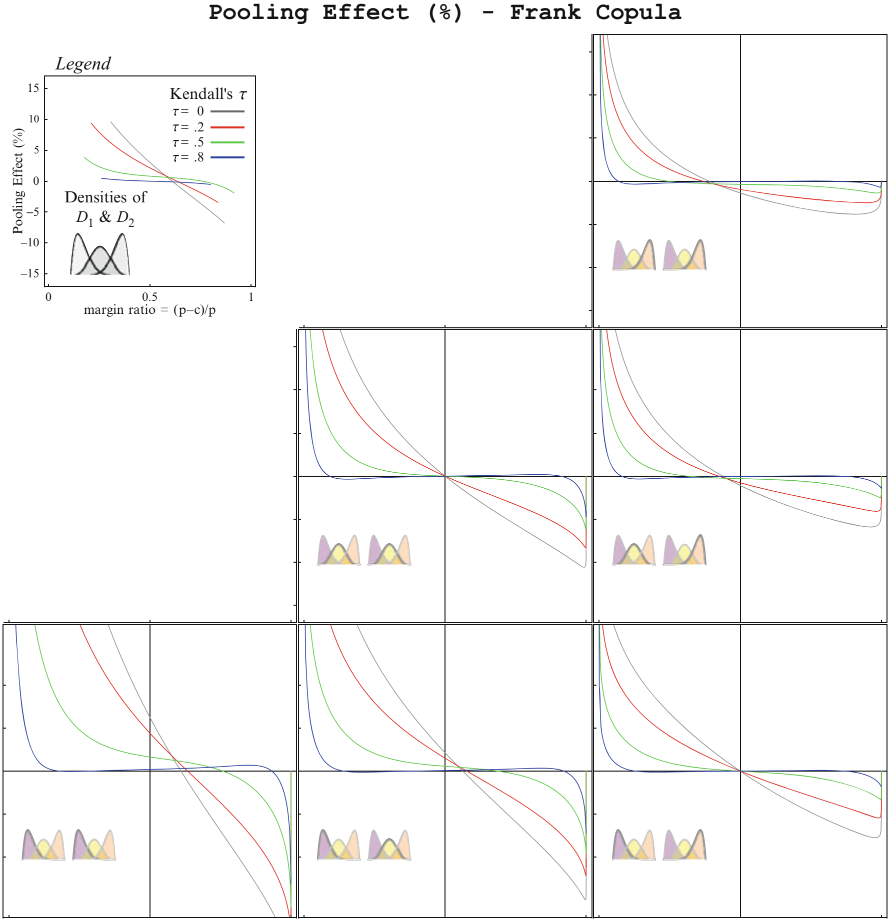


Fig. 3.6 The plots of percentage pooling effect on inventory levels under Frank copula with positive Kendall's τ. The marginal distributions are, from *top down*, $\beta(2,8)$, $\beta(5,5)$, and $\beta(8,2)$, and from *left to right*, $\beta(2,8)$, $\beta(5,5)$, and $\beta(8,2)$. These represent left-skewed, symmetrical, and right-skewed marginals. The common legend for subplots is given on the *top-left corner*

We start with some general conclusions that can be drawn from Figs. 3.4–3.7. First, a stronger dependence measured by a high Kendall's τ leads to a pooling effect that is smaller in absolute value. This is expected, as we know that for co-monotone demand distributions, the sum of quantiles of individual demand distributions is equal to the quantile of the sum of the two demand distributions, implying no pooling effect.

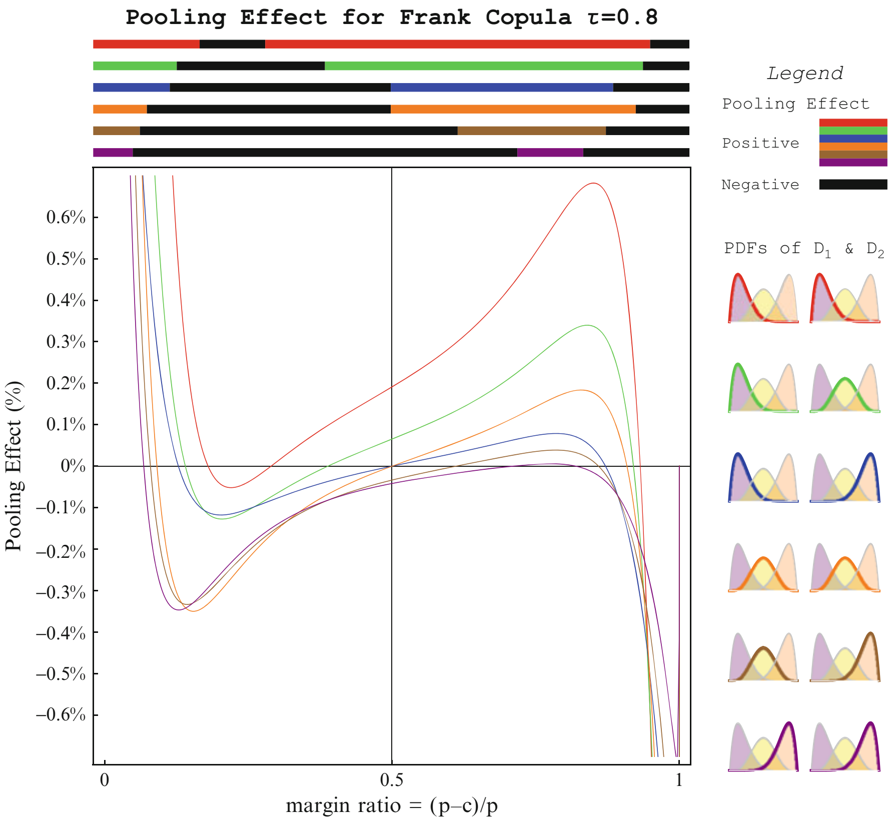


Fig. 3.7 The plots of percentage pooling effect on inventory levels under Frank copula with negative Kendall's τ. The marginal distributions are, from *top down*, $\beta(2,8)$, $\beta(5,5)$, and $\beta(8,2)$, and from *left to right*, $\beta(2,8)$, $\beta(5,5)$, and $\beta(8,2)$. These represent left-skewed, symmetrical, and right-skewed marginals. The common legend for subplots is given on the *top-left corner*

3.5.2.1 Gumbel Copula

For the high-dependence case ($\tau = 0.8$), we see that pooling does not have a strong effect on inventory levels for most margin ratios. This is consistent with previous knowledge. One exception is when the margin ratio is very low: for any marginal density combination, low margin ratios result in positive pooling effect. Another observation is that as the dependence decreases, the threshold value increases, implying a positive pooling effect for even smaller values of margin ratio when dependence is high.

3.5.2.2 Clayton Copula

Compared to Gumbel copula, we find two important differences. First, the Clayton copula implies a higher threshold value t_0 compared to Gumbel, given everything the same. Second, this threshold value increases as the dependence increases. This observation is in stark contrast to the one with Gumbel copula. Recall that Clayton copula shifts density toward the left tail of the joint distribution, which is the underlying reason behind these differences.

3.5.2.3 Frank Copula

This copula can cover both the positive and negative dependence cases. The threshold value t_0 is more robust to the skewness of the marginals as compared to the first two copulas. This is due to fact that Frank copula's tail dependencies are symmetric. The effect of this symmetry can also be seen across the pooling effect lines at different τ levels.

One case deserves a detailed discussion. When Kendall's $\tau = 0.8$, we find that the threshold value is not unique. In Fig. 3.8, we plot the percentile pooling effect for all six marginal distribution combinations. On the top of the figure, we plot the regions of the pooling effect where it is positive or negative. For any combination, we find that pooling requires higher inventory levels in two different disjoint regions of the margin ratio. Hence, the uniqueness of the threshold value is not valid for this particular copula when there is very high dependence. We should point out that the magnitude of pooling effect is quite small in this case, since co-monotonicity leads to no pooling effect.

With the Frank copula, one can model negative correlation which provides additional insights. As expected, negative correlation can lead to significant pooling effects, especially when the margin ratio is high or low. When demands are highly negatively correlated, the pooled inventory level (which depends on the sum of these two random variables) is robust against the margin ratio. However, the dedicated inventory levels are small for low margin ratios, and high for high margin ratios. Hence, we see that pooling effect is significantly positive for lower ratios and significantly negative for high margin ratios.

3.5.2.4 Kendall's τ Versus Pooling Effect

Finally, we address how the magnitude of the pooling effect varies with model parameters. In particular, we investigate whether the optimal pooled inventory level varies monotonically with the dependence. Under normality, we know that it does. Under general dependence structures, however, monotonicity of pooled inventory

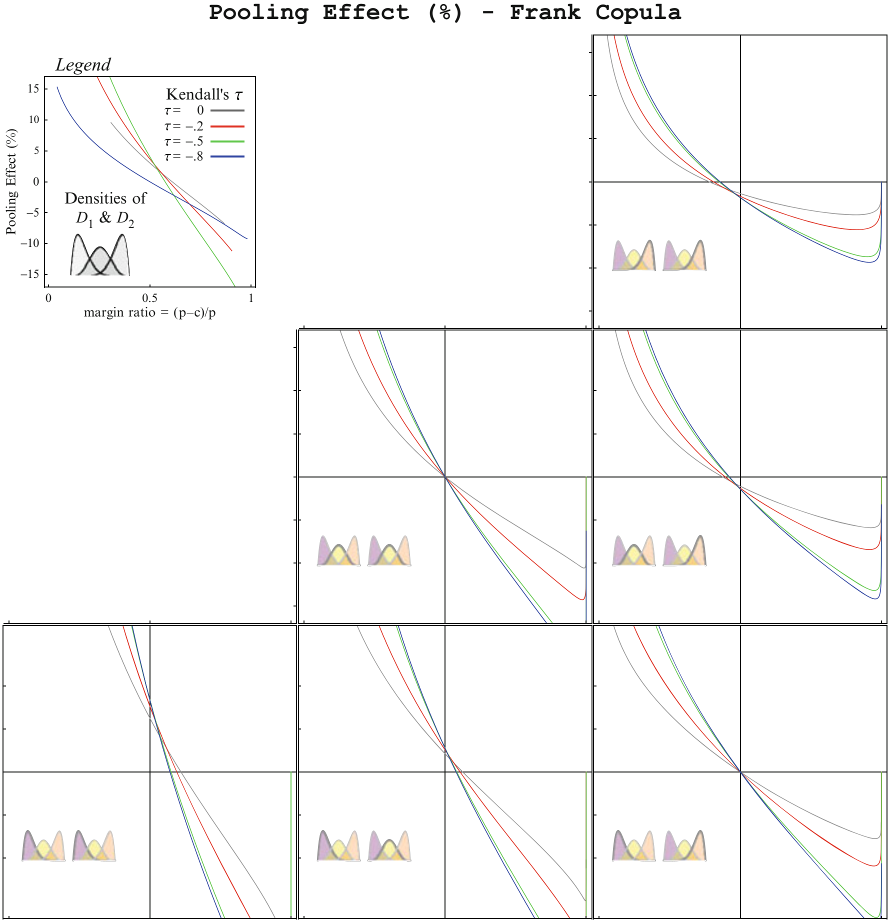


Fig. 3.8 The plots of percentage pooling effect on inventory levels under Frank copula with Kendall's $\tau = 0.8$. The pooling effect for different marginal distribution combinations, $\beta(2,8)$, $\beta(5,5)$, and $\beta(8,2)$, are depicted in different colors. *Top figure* shows the regions where pooling effect is negative or positive

level with respect to dependence does not hold. Figure 3.9 presents the pooled inventory levels under different copulas with identical normal marginals compared to dedicated inventory levels, under three different margin ratios.

Figure 3.9 clearly shows the importance of margin ratio. For a low margin ratio, we see that all three copulas we tried show higher inventory levels compared to dedicated. For high margins, the opposite seems to be the most common trend. For mid-range margin ratios, Clayton copula shows the most variation as the Kendall's τ changes.

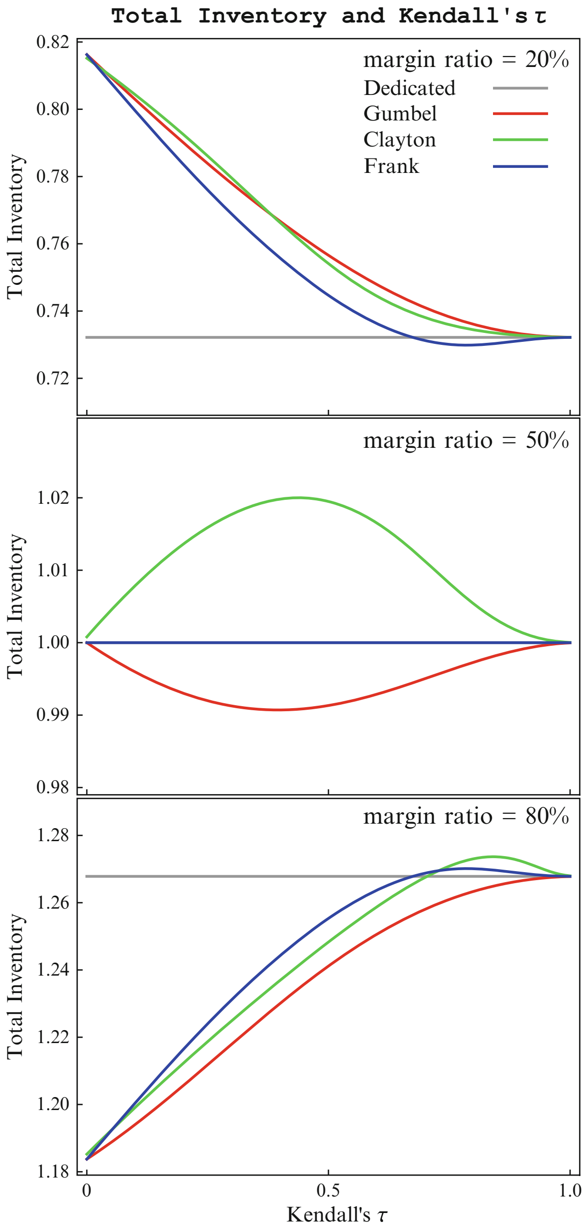


Fig. 3.9 The total inventory levels under three different margin ratios, varying with respect to dependence (Kendall's τ). *Top panel* gives the low margin case with $t = 0.2$, *middle panel* is the medium margin case with $t = 0.5$, and the *bottom panel* is the high margin case with $t = 0.8$. The *gray line* indicates dedicated inventory level, *red* is the pooled inventory level under Gumbel copula, *green* is the pooled inventory level under Clayton copula and *blue line* is the pooled inventory level under Frank copula. The marginals are $\beta(5,5)$

3.6 Conclusion

We investigate the optimal inventory levels after pooling to determine whether the manager should increase inventory levels after a switch to pooling. We show that one has to understand the true underlying dependency structure between the individual demand sources, as well as the uncertainty within each demand source, to determine the right level of inventories. In order to understand the effect of each factor, we use copula theory to separate the effect of demand source uncertainty and dependencies between these demand sources, and study the interactions in between, as well as the effects of these interactions on the inventory levels.

We find that the sign of pooling effect depends on the margin ratio: It is positive if and only if margin ratio is higher than a threshold, under certain conditions. This threshold depends on the marginal demand distributions as well as the copula that joins them. Through numerical studies, we investigate these relationships and conclude that tail dependencies and the strength of dependency are the main factors that affect this threshold. Finally we show that pooled inventory levels are not necessarily monotone with respect to the level of dependence. This is especially true for copulas with asymmetric tail dependencies.

There are open questions that requires further research. In this paper, we focus on the unimodal distributions. If the marginals of the demand distributions are bimodal, then pooling effect is expected to be stronger around these modes, especially if these modes are closer. We also assume that one can estimate the true copula structure underlying the data. When the quality of this estimation is not high or it is not available at all, one will need to consider all possible dependence structures given the limited information, and come up with upper and lower bounds of true optimal inventory levels. Incorrect estimation of the dependence might also cause setting incorrect inventory levels.

In this paper we consider the case with two products. However, the copula framework is able to handle any number of marginals. Therefore it is possible to easily extend the results of this paper to case with arbitrary finite number of products. Our conjecture is that, when the marginals are identical, the results of the multi-item case will be similar to our results when the two marginals are identical. When they are not identical however, the relative shapes of marginals is critical in determining the inventory levels. We leave these questions to future work.

Product characteristics are critical in determining pooled inventory levels. This paper covers the case with perfectly substitutable products. When products are not perfectly substitutable, then pooling effect will be smaller. At the extreme case when pooled products are uniquely different from each other, no pooling will be possible. Hence a deeper understanding of moderate levels of substitutability is required to investigate those situations. Finally, another assumption in this paper is that the products are financially identical: They have the same revenue and cost per unit. When this assumption is not valid, one will need to have a fulfilment policy as to which demand source to satisfy when there is insufficient inventory. This policy structure would determine the pooling inventory levels and in turn the efficiency of pooling.

Acknowledgements We would like to thank the two anonymous referees for their helpful comments.

References

Alfaro, J. A., & Corbett, C. J. (2003). The value of SKU rationalization in practice (the pooling effect under suboptimal inventory policies and nonnormal demand). *Production Operations Management*, *12*, 12–29.

Corbett, C. J., & Rajaram, K. (2006). A generalization of the inventory pooling effect to nonnormal dependent demand. *Manufacturing and Service Operations Management*, *8*(4), 551–358.

Eppen, G. D. (1979). Effects of centralization on expected costs in a multi-location newsboy problem. *Management Science*, *25*(5), 498–501.

Erkip, N., Housman, W. H., & Nahmias, S. (1990). Optimal centralized ordering policies in multi-echelon production and inventory problems. *Management Science*, *36*(3), 381–392.

Gerchak, Y., & Mossman, D. (1992). On the effect of demand randomness on inventory and costs. *Operations Research*, *40*, 804–807

Gerchak, Y., & He, Q. M. (2003). On the relation between the benefits of risk pooling and the variability of demand. *IIE Transactions*, *35*, 1027–1031.

Jang, J., & Jho, J. H. (2007). Asymptotic super(sub)additivity of value-atRisk of regularly varying dependent variables. Preprint MacQuarie University, Sydney.

Kaas, R., Laeven, R. J. A., & Nelsen, R. B. (2009). Worst VAR scenarios with given marginals and measures of association. *Insurance: Mathematics and Economics*, *44*, 146–158.

Liu, J., & David, H. A. (1989). Quantiles of sums and expected values of ordered sums. *Australian Journal of Statistics*, *31*(3), 469–474.

McNeil, A. J. and Frey, R., & Embrechts, P. (2005). Quantitative risk management: concepts, techniques, tools. Princeton Series in Finance.

Nelsen, R. R. (1999). *An introduction to copulas*. New York: Springer-Verlag.

Netessine, S., & Rudi, N. (2003). Centralized and competitive inventory models with demand substitution. *Operations Research*, *51*(2), 329–335.

Pasternack, B., & Drezner, Z. (1991). Optimal inventory policies for substitutable commodities with stochastic demand. *Naval Research Logistics*, *38*, 221–240.

Van Mieghem, J. A., & Rudi, N. (2002). Newsvendor networks: dynamic inventory management and capacity investment with discretionary pooling. *Manufacturing Service Operations Management*, *4*(4), 313–335.

Yang, H., & Schrage, L. (2009). Conditions that cause risk pooling to increase inventory. *European Journal of Operational Research*, *192*, 837–851.

Chapter 4
Repeated Newsvendor Game with Transshipments

Xiao Huang and Greys Sošić

Abstract We study a repeated newsvendor game with transshipments. In every period n, retailers face a stochastic demand for an identical product and independently place their inventory orders before demand realization. After observing the actual demand, each retailer decides how much of her leftover inventory or unsatisfied demand she wants to share with the other retailers. Residual inventories are then transshipped in order to meet residual demands, and dual allocations are used to distribute residual profit. Unsold inventories are salvaged at the end of the period. While in a single-shot game retailers in an equilibrium withhold their residuals, we show that it is a subgame-perfect Nash equilibrium for the retailers to share all of the residuals when the discount factor is large enough and the game is repeated infinitely many times. We also study asymptotic behavior of the retailers' order quantities and discount factors when n is large. Finally, we provide conditions under which a system-optimal solution can be achieved in a game with n retailers, and develop a contract for achieving a system-optimal outcome when these conditions are not satisfied. This chapter is based on Huang and Sošić (European Journal of Operational Research 204(2):274–284, 2010).

Keywords Transshipment • Subgame perfect Nash equilibrium • Repeated game • Asymptotic behavior

---

X. Huang
John Molson School of Business, Concordia University, Montréal, Québec, Canada, H3G 1M8
e-mail: xiaoh@jmsb.concordia.ca

G. Sošić (✉)
Marshall School of Business, University of Southern California, Los Angeles, CA 90089, USA
e-mail: sosic@marshall.usc.edu

T.-M. Choi (ed.), *Handbook of Newsvendor Problems*, International Series in Operations Research & Management Science 176, DOI 10.1007/978-1-4614-3600-3_4,

4.1 Introduction and Literature Review

As the intensity of the business competition grows, retailers and distributors want to achieve more flexibility and become more responsive to their customers. However, fulfilling the demand is a challenge given the high uncertainty of the market, the limited capacity, and the tight budget constraints. In many such situations, it is worthwhile for the distributors to form alliances that will share substitutable inventory or services. Such cooperation is even more beneficial when products have short life/sales cycles, become obsolete fast, face long suppliers' lead times, and customer's demands that are hard to predict. Examples of such products are apparel, pop music, and high-tech products, among others.

Many retail chains implement transshipments or inventory sharing (we will use both terms in this article) among their stores. For example, Takashimaya, a Japanese department store chain, adopts inventory sharing policies among its stores by allowing sales persons to search on their PDAs the inventories held by other branches when the product is not held in stock at their location. The requested product is received the following day. In this way, Takashimaya manages to optimize the inventory within specialized shops. Similar policies are implemented in Music Millenium, Guess, and others.

While inventory sharing within a company is, intuitively, feasible and profitable, it is worthwhile to mention that similar practices happen among independent parties as well. iSuppli.com markets itself as the "collaborative ground" and is trying to build up a network of unrelated parties that need the same electronic components.

When inventory sharing is introduced into the system, various questions need to be addressed:

1. *Inventory Decision*: One of the merits of inventory sharing is the reduction of the overstocking cost, because inventory-sharing parties usually hold less inventories.[1] An important question here is, to what extent are the inventory positions going to be reduced?
2. *Transshipment*: When multiple retailers participate in transshipments, how to allocate the inventory among them? The transshipping pattern can be either determined a priori by a contract (i.e., the retailer with surplus inventory may select where her inventory is going), or a posteriori according to some objective (i.e., maximize the total profit of all retailers).
3. *Profit Allocation*: How are the profits generated from transshipments allocated among the retailers? For example, there may be a flat-rate price for each unit transshipped, or the total profit can be divided evenly among all participating retailers.
4. *Sharing Decision*: How much of their leftover inventories or unsatisfied demands are the retailers willing to share with others? Are they going to put all their

[1]For some exceptions, see Yang and Schrage (2009), which show that the inventory levels can increase after centralization when demand follows right-skewed distributions, or when the newsvendor ratio is low.

leftovers (inventories or demands) on the table, or strategically withhold some of them? This decision may depend upon initial inventory position, transshipment policies, or profit allocation.

5. *Time Horizon*: Is inventory sharing a one-time event, or an activity in which the retailers will be engaged repeatedly? In the latter case, are the unsold inventories carried over to the next period, or salvaged at the end of the period?

Many of these questions have been addresses by the researchers in various combinations, as inventory sharing has been a subject of extensive research work. One stream of research focuses on inventory decisions. Parlar (1988) develops a game-theoretical model for substitutable products in which leftover inventory and unmet demand are matched through customer-driven search. This implicitly means that the party that holds the excess inventory receives the entire profit from inventory sharing. The paper proves that a first-best outcome (that is, a system-optimal solution) can be achieved in a two-retailer game. Wang and Parlar (1994) analyze a similar problem with three retailers. They find that the core of the game can be empty, and thus inventory sharing between sub-coalitions of players may occur. Lippman and McCardle (1997) consider an environment with aggregated stochastic industry demand, which has to be divided among different firms. They study the relationship between initial demand-sharing rules and equilibrium inventory decisions, and they determine conditions for a unique equilibrium.

Another stream of research analyzes transshipment of inventories. Among more recent papers, Dong and Rudi (2004) examine the impact of horizontal transshipments between the retailers on both the retailers and on the manufacturer, while Zhang (2005) generalizes their results. Rudi et al. (2001) and Hu et al. (2007) study decision making in decentralized systems and the significance of transshipment prices in local decisions. Wee and Dada (2005) consider a one-warehouse n-retailer system in which the retailers can receive inventory from the warehouse and from the other retailers. They analyze the impact of the number of retailers and demand correlation on transshipment decisions. Zhao et al. (2005) study a model in which the retailers determine both a base-stock policy (for inventory stocking) and a threshold policy (for inventory sharing) prior to demand realization. Shao et al. (2011) study a supply chain which is both vertically and horizontally decentralized. They show the importance of the transshipment price in determining whether firms benefit or lose from transshipment, and investigate how the control of the parameters of the transshipment decision affects firms' transshipment incentives.

If the retailers agree to share their residuals, a decision has to be made as to how to allocate the additional profit generated through transshipment of inventories. This decision can be made jointly by the retailers, or it can be, for instance, chosen by a manufacturer whose products they are selling, or a trade association, or a larger organization to which the retailers belong. Clearly, different allocation rules will have different impacts on the retailers' stocking quantities, on the amount of inventories shared among retailers, and on the profit levels realized in the system. Ideally, the retailers would want to choose an allocation rule that would maximize the additional profit from transshipments. In order to achieve this goal, it is sufficient that the allocation rule:

(a) Induces participation of *all* retailers.
(b) Motivates the retailers to share *all of their residuals* with others.

We will call condition (a) the *full participation* condition, and condition (b) the *complete sharing* condition. Anupindi et al. (ABZ 2001) and Granot and Sošić (G&S 2003) develop a multistage model for a problem in which n independent retailers face stochastic demands for identical products. In the first stage, before the demand is realized, retailers unilaterally determine their stocking quantities. After the demand is realized and the retailers fulfill their own demands with inventories on hand, some retailers are left with unsatisfied demand, while others have leftover supply. The retailers at this point cooperatively determine a transshipment pattern for distribution of residual inventories among themselves. The additional profit generated through transshipments (which we call *residual profit*) is divided according to an allocation rule specified at the beginning of the game. ABZ formulate a two-stage model for this problem and implicitly assume that the retailers share all of their residuals with the others. Thus, the complete sharing condition is automatically satisfied. They propose a core allocation rule based on the dual prices for the transshipment problem (referred hereafter as *dual allocations*; for detailed description, see Sect. 4.2), which satisfies the full participation condition. ABZ point out that dual allocations, in general, do not induce a first-best solution. When the retailers are allowed to withhold some of their residuals, G&S show that dual allocations may not be able to induce complete sharing of the residual supply/demand. This may, in turn, reduce the residual profit. On the other hand, monotonic allocation rules (such as the fractional rule and the Shapley value) satisfy the complete sharing condition, but these rules, in general, do not belong to the core, and thus they violate the full participation condition. Consequently, some retailers may form inventory sharing subcoalitions, which, in turn, may result in a reduced residual profit. Notice that all of the above conclusions hold in a myopic framework. If the retailers are farsighted and consider possible further reactions of their inventory-sharing partners to their actions, Sošić (2004) shows that complete inventory sharing among all retailers is a stable outcome when the residual profit is distributed according to the Shapely value allocations.

In this work, we study the extension of the above one-shot game from G&S to a repeated setting, in which each retailer faces her demand over several periods. In each period, the three-stage model corresponds to that described in the one-shot game. We want to point out that we are interested in studying the impact of the repeated interactions on the retailer's decisions in the second stage (how much of their residuals they want to share with others) and on selecting their partners for inventory-sharing (possible formation of subcoalitions). As a result, we continue to assume the newsvendor framework, in which unsold inventories are salvaged at the end of each period and no demand is backlogged. This setting is common, for example, for fashion goods or high-tech items. In addition, we assume that the retailers in each period sell a product with *identical* characteristics (demand distribution, cost, and price). This is a simplifying assumption, which nevertheless may approximate many real-life situations, in which items with *similar* characteristics are sold in different periods. For instance, every season apparel manufacturers introduce new collections. One can presume that items that fall into same categories (t-shirts or other casual clothing, business suits, or trendy items made by the same company) will have similar demand characteristics in different years. A similar conclusion can be made for Christmas toys (say, different versions of Barbie or Elmo dolls), music (new CDs released by Prince, Lady Gaga, or Carrie Underwood), etc. In the high-tech industry, new hard disk drives or new processors are introduced on a regular basis to replace the previous generation of corresponding products. As the technology advances and the models with better performance reach the market, one can assume that the new product will have demand and price similar to the original demand and price of the product that it is replacing. Note that our model also covers some instances in which the prices change in different periods—we discuss this in more detail in Sect. 4.3.

As mentioned earlier, when the retailers cooperatively generate additional profit, they have to decide how to distribute it among themselves. In our model, we assume that the retailers apportion this extra income according to the *dual allocations*. These allocations are based on the dual solution of the linear programming problem (4.1) used to determine the optimal shipping pattern for residuals, and are, therefore, easy to compute. For detailed description of the model and dual allocations, please see Sect. 4.2. As shown by ABZ, dual allocations are in the core of the corresponding game, which makes the coalition of all players stable, because no players (or subsets of players) benefit from a defection, and hence dual allocations satisfy the full participation condition. Thus, if each retailer shares all of her residuals, the profit from inventory sharing is maximized. However, if players are allowed to withhold some of their residuals, G&S show that players will not share all of their leftover inventory/unmet demand, which, in turn, reduces the profit obtained through inventory sharing. Note, however, that these results hold in a one-shot setting, where players do not consider future interactions. Now, in a repeated game, we want to address the following questions:

1. When the retailers interact repeatedly, what is the impact of the length of the time horizon on the retailers' decisions, and is it possible to induce the retailers to share all of their residuals with dual allocations?
2. Under what conditions can a first-best solution be achieved without additional enforcement mechanisms, and what type of contracts can induce system-optimal decisions when these conditions do not hold?

The answers to the first question are obtained through standard game-theoretical tools. We show that dual allocations induce the retailers to withhold residuals when the game is played a finite number of times. On the other hand, the retailers in the infinite-horizon model may be induced to share all of their residuals when they put enough weight on their future payoffs. As the number of retailers increases, calculation of the lower bound for the value of the discount factor that induces the complete sharing of residuals becomes intractable. However, we are able to obtain some asymptotic results for a large number of players. We also demonstrate that a complete sharing of residuals may be induced when the punishment (that is, nonsharing of inventories) is not enforced over an infinite horizon.

In answering the second question, we provide a condition for achieving a first-best outcome, and develop a contract that leads to a first-best outcome when some retailers' optimal stocking decisions differ from the system-optimal ones.

The structure of this article is as follows: we briefly introduce the one-shot inventory sharing game in Sect. 4.2, and in Sect. 4.3 we extend this model to a repeated setting. In Sect. 4.4, we develop some asymptotic results for the retailers' ordering quantities and lower bounds on discount factors that induce complete sharing of residuals for large number of players. In Sect. 4.5, we derive conditions for achieving a first-best outcome without additional enforcement mechanisms, while in Sect. 4.6 we develop a contract that induces a first-best solution in a more general setting. We conclude in Sect. 4.7. Longer proofs are given in a technical appendix.

4.2 One-Shot Inventory-Sharing Game

Each period in our repeated game corresponds to the three-stage inventory-sharing model from G&S and can be described as follows. We use $N = \{1, 2, \ldots, n\}$ to denote a set of retailers who are selling an identical product. We assume that the retailers face independent random demands, D_i, and that each retailer knows the distribution of her demand, F_i, and its density, f_i. After demands are realized and each retailer satisfies her own demand from inventory on hand, the retailers can share their residuals—leftover inventories or unsatisfied demands. The total profit from transshipments—residual profit—has to be divided among the retailers according to an allocation rule agreed upon by all of them before the game begins. We assume that there are no capacity constraints and that the game begins with zero inventory. The three stages are modeled as follows:

Stage 1: Before demand D_i is realized, each retailer independently makes her own ordering decision, X_i, contingent upon the demand distribution and the allocation rule that will be used to distribute the residual profit.

Stage 2: After demand is realized, each retailer decides how much of her residuals she would like to share with others. Let $\bar{H}_i = \max\{X_i - D_i, 0\}$ and $\bar{E}_i = \max\{D_i - X_i, 0\}$ denote the total leftover inventory and unsatisfied demand for retailer i, respectively. We will use bold letters to denote vectors, that is, $(\bar{\mathbf{H}}, \bar{\mathbf{E}}) = (\bar{H}_1, \ldots, \bar{H}_n, \bar{E}_1, \ldots, \bar{E}_n)$. We denote the retailers' sharing decisions (amounts of residual supply/demand that retailer i decides to share with the other retailers) by H_i and E_i, respectively. It is straightforward that H_i and E_i must satisfy $0 \le H_i \le \bar{H}_i$, $0 \le E_i \le \bar{E}_i$.

Stage 3: The shipping pattern for leftover inventory that maximizes the residual profit is determined. The resulting residual profit is then distributed among the retailers according to the allocation rule determined before the first stage takes place (in this article, we assume that the retailers use dual allocations). Any inventory left at the retailers is salvaged.

Let r_i, c_i, and v_i denote, respectively, the unit retail price, cost, and salvage value for retailer i, Y_{ij} and t_{ij} denote the amount of stock shipped and the unit cost of transshipment from retailer i to retailer j. We assume that $r_i > r_j - t_{ij}$, that is, each retailer satisfies her own demand first, and $v_i - t_{ji} < v_j$, that is, the retailers do not benefit from salvaging unsold items at other locations.

We next present some results from G&S (2003). The transshipment pattern in the third stage, given demand realizations and retailers' sharing decisions, can be solved through linear programming. Let $R(\mathbf{X},\mathbf{D},\mathbf{H},\mathbf{E})$ denote the residual profit from the transshipments; as the retailers sharing decisions, $(\mathbf{H},\mathbf{E})$, depend on the actual residual values, $(\bar{\mathbf{H}},\bar{\mathbf{E}})$, this profit is a function of the retailers' order sizes and demand realizations, $\mathbf{X}$ and $\mathbf{D}$. The optimal shipping pattern, $R^*(\mathbf{X},\mathbf{D},\mathbf{H},\mathbf{E})$, can be determined by solving the following linear programming problem.

$$R^*(\mathbf{X},\mathbf{D},\mathbf{H},\mathbf{E}) := \max_{\mathbf{Y}} \sum_{i,j=1}^{n} (r_j - v_i - t_{ij})Y_{ij}, \tag{4.1a}$$

$$\text{subject to: } \textstyle\sum_{j=1}^{n} Y_{ij} \le H_i \quad i = 1,2,\ldots,n, \tag{4.1b}$$

$$\textstyle\sum_{j=1}^{n} Y_{ji} \le E_i \quad i = 1,2,\ldots,n, \tag{4.1c}$$

$$Y_{ij} \ge 0 \qquad i,j = 1,2,\ldots,n. \tag{4.1d}$$

We denote the allocation of residual profit to retailer i by $\varphi_i^d(\mathbf{X},\mathbf{D},\mathbf{H},\mathbf{E})$. If λ_i and μ_i denote the dual prices corresponding to the constraints (4.1b) and (4.1c), respectively, then $\varphi_i^d(\mathbf{X},\mathbf{D},\mathbf{H},\mathbf{E}) = \lambda_i H_i + \mu_i E_i$, and the profit for a retailer, i, can be written as:

$$P_i^d(\mathbf{X},\mathbf{D},\mathbf{H},\mathbf{E}) = r_i \min\{X_i, D_i\} + v_i \bar{H}_i - c_i X_i + \varphi_i^d(\mathbf{X},\mathbf{D},\mathbf{H},\mathbf{E}).$$

Given the stocking quantity decisions and demand realizations, $\mathbf{X}$ and $\mathbf{D}$, the retailers in the second stage of the game make their sharing decisions according to the Nash equilibrium (NE), which we denote by $(\mathbf{H}^{\mathbf{X},\mathbf{D}},\mathbf{E}^{\mathbf{X},\mathbf{D}})$. Thus, they must satisfy the following inequalities:

$$P_i^d(\mathbf{X},\mathbf{D},\mathbf{H}^{\mathbf{X},\mathbf{D}},\mathbf{E}^{\mathbf{X},\mathbf{D}}) \ge P_i^d(\mathbf{X},\mathbf{D},H_i,H_{-i}^{\mathbf{X},\mathbf{D}},E_i,E_{-i}^{\mathbf{X},\mathbf{D}}),$$
$$\forall H_i \le \bar{H}_i,\ E_i \le \bar{E}_i,\ i = 1,2,\ldots,n,$$

where $x_{-i} = (x_1, x_2, \ldots, x_{i-1}, x_{i+1}, \ldots, x_n)$.

Finally, the first-stage NE ordering decisions, $\mathbf{X}^{\mathrm{d}}$, must satisfy

$$J_i^{\mathrm{d}}(\mathbf{X}^{\mathrm{d}}) \ge J_i^{\mathrm{d}}(X_i, \mathbf{X}_{-i}^{d}),$$

where $J_i^{\mathrm{d}}(\mathbf{X}) = \mathsf{E}[P_i^d(\mathbf{X},\mathbf{D},\mathbf{H}^{\mathbf{X},\mathbf{D}},\mathbf{E}^{\mathbf{X},\mathbf{D}})]$ is retailer i's expected profit when retailers' ordering decisions form vector $\mathbf{X}$. Huang and Sošić (2010a) provide conditions for existence of the NE in ordering quantities, $\mathbf{X}^{\mathrm{d}}$, for this game. As we are primarily interested in the effects of repeated interactions on players' decisions, in what follows we assume that these conditions are satisfied and that the NE exists.

We also mention, as benchmarks, two related models—the game without transshipments and the centralized model. If the retailers do not share their residuals, each retailer's profit can be described as

$$P_i^1(X_i, D_i) = r_i \min\{X_i, D_i\} + v_i \bar{H}_i - c_i X_i,$$

with the expectation $J_i^1(X_i) = \mathsf{E}[P_i^1(X_i, D_i)]$. Superscript 1 denotes the model in which each retailer acts individually. The optimal ordering decision, X_i^1, corresponds to the newsvendor solution. In the centralized model, in which a single decision maker optimizes the profit of the entire system, the total system profit can be written as

$$P^n(\mathbf{X}, \mathbf{D}) = \sum_{i=1}^{n} r_i \min\{X_i, D_i\} + v_i \bar{H}_i - c_i X_i + \bar{R}^*(\mathbf{X}, \mathbf{D}),$$

with the expectation $J^n(\mathbf{X}) = \mathsf{E}[P^n(\mathbf{X}, \mathbf{D})]$. Superscript n denotes that n retailers participate in inventory sharing. The optimal ordering amount for this model, $\mathbf{X}^n$, maximizes the total system profit and is referred to as a *first-best solution*.

4.3 Repeated Inventory-Sharing Game

In this section, we study the inventory-sharing game in a repeated setting. When the retailers do not expect future interactions with their inventory-sharing partners, dual allocations preclude them from formation of subcoalitions, but may also provide an incentive for some (or all) of them to withhold a portion of their residuals (which may increase their allocations). The main topic of our interest is to study the impact of repeated interactions on the retailers' sharing decisions in the second stage. Our repeated game is modeled identically in every period, following the steps described in the one-shot model. The goal of each retailer is to maximize her total discounted profit, and we consider both a finite and an infinite horizon. A solution concept commonly used in this setting is *subgame perfect Nash equilibrium* (SPNE)—a solution in which players' strategies constitute a NE in every subgame of the original game.

We assume that unsold inventories are salvaged at the end of each period and that inventory level at the beginning of each period is zero. If we allow the retailers to strategically increase their orders in one period and transfer a portion of inventory to the next period, the result would be a significantly more complicated model that is beyond the scope of this work. In addition, when making her decision, each retailer knows the entire history of previous decisions for all retailers. While this assumption may be rather strong, it is not uncommon in the repeated-game setting to assume that all players know the entire history (see, for instance, Bagwell and Staiger 1997;

Haltiwanger and Harrington 1991; Rotemberg and Saloner 1986). We feel that such an assumption may be appropriate, say, for settings in which the retailers belong to a larger organization, or within a trade association.

G&S (2003) show that the retailers who share inventory only once withhold some of their residuals. By using standard game-theoretical tools, it can be easily shown that the same is true when the game is repeated a finite number of times; hence, we state our next result without a proof.

Proposition 1. *Complete sharing is not achieved if the inventory-sharing game with n retailers is repeated a finite number of times.*

We next consider an infinitely repeated game and introduce the *Nash reversion* strategy (NRS), which can be described as follows: each retailer completely shares her residuals until one or more of them deviate by withholding some of their residuals. From that moment on, no residuals are shared in the subsequent periods by any of the retailers. We show that this strategy is an SPNE.

Let P_{it} and X_{it} denote the profit and the ordering quantity of retailer i in period t, respectively; we use similar notation for her shared and actual residuals in period t, H_{it}, E_{it} and $\bar{H}_{it}$, $\bar{E}_{it}$. The retailers' decisions are based on previous histories, $h_{t-1} = \{\mathbf{X}_l, \mathbf{H}_l, \mathbf{E}_l\}_{l=1}^{t-1}$, that include stocking quantities and shared residuals in all periods preceding t. Thus, we write $(X_{it}, E_{it}, H_{it})(h_{t-1})$ to denote that (X_{it}, E_{it}, H_{it}) depends on h_{t-1}. We let $(h_{t-1})_l = (\mathbf{X}_l, \mathbf{H}_l, \mathbf{E}_l)$ denote the retailers' decision in period l. Recall that X_i^{d} and X_i^{I} denote the optimal stocking quantities in one-shot games with dual allocations and without transshipments, respectively, and that δ denotes the discount factor. The following result can be shown through the application of the folk theorem.

Theorem 1. *Suppose that an inventory sharing game with n retailers is repeated infinitely many times. Then, there exists* $\delta_n^* \in (0,1)$ *such that the NRS, in which*

$$(X_{it}, E_{it}, H_{it})(h_{t-1}) = \begin{cases} (X_i^{\mathrm{d}}, \bar{H}_{it}, \bar{E}_{it}) & \text{if } t = 1 \text{ or } (h_{t-1})_l = (\mathbf{X}^{\mathrm{d}}, \bar{\mathbf{H}}, \bar{\mathbf{E}}), \\ & \forall l = 1, \ldots, t-1 \\ (X_i^{\mathrm{I}}, 0, 0) & \text{otherwise,} \end{cases}$$

constitutes a SPNE of the infinitely repeated game whenever $\delta > \delta_n^*$.

The lower bound for the discount factor, δ_n^*, can in practice be difficult to evaluate, so in Sect. 4.4, we explore in more detail its asymptotic behavior. We illustrate our result with the following numerical example.

Example 1. Suppose that $n = 3$, all three retailers face two-point demand which can achieve 0 with probability 0.5 and 10 with probability 0.5, and c_i=3.7; r_i=10; $v_i = 1$, $i = 1,2,3$; $t_{ij} = 1$, $i, j = 1,2,3, i \neq j$. When the retailers share their inventory and distribute the residual profit according to dual allocations, their individual stocking quantities decrease from 10 to 7, and the corresponding expected profits increase from 18 to 22. The discount factors that induce complete residual sharing by all retailers satisfy $\delta > \delta_3^* = 0.93$.

4.3.1 Finite Punishment Period

The NRS represents the belief that "once the trust is lost, it is lost forever." However, one can object that infinite punishment may not be credible, because besides punishing the defecting retailer, it hurts the punishers as well. Hence, we consider a "milder" strategy in which punishment lasts only for a finite number of periods before the retailers recover from the "bad memories" and return to cooperation. In this framework, only the history of the past k periods, $h_{t-k}^{t-1} = \{\mathbf{X}_\tau, \mathbf{H}_\tau, \mathbf{E}_\tau\}_{\tau=t-k}^{t-1}$, has an impact on retailers' decisions.

Theorem 2. *Suppose that an inventory sharing game with n retailers is repeated infinitely many times. Then, there exists $k_n^* \in \mathbf{N}$ such that $\forall k > k_n^*$ there is a $\delta_n^*(k)$ such that the strategy in which*

$$(X_{it}, E_{it}, H_{it})(h_{t-k}^{t-1}) = \begin{cases} (X_i^{\mathrm{d}}, \bar{H}_{it}, \bar{E}_{it}) & \text{if } t=1 \text{ or } (h_{t-k}^{t-1})_\tau = (\mathbf{X}^{\mathbf{1}}, \mathbf{0}, \mathbf{0}) \ \forall \tau = 1, \ldots, k \\ & \text{or } (h_{t-k}^{t-1})_{t-1} = (\mathbf{X}^{\mathrm{d}}, \bar{\mathbf{H}}, \bar{\mathbf{E}}) \\ (X_i^1, 0, 0) & \text{otherwise,} \end{cases}$$

constitutes an SPNE of the infinitely repeated game whenever $\delta > \delta_n^*(k)$.

The proof is again obtained through the application of the folk theorem. If a player, j, considers a deviation from $(X_j^{\mathrm{d}}, \bar{H}_{jt}, \bar{E}_{jt})$, any momentary gain is canceled by future reduction in payoffs when the discount factor is large enough and the punishment is carried over an appropriate number of periods. During the punishment period, each retailer plays her optimal strategy for noncooperative setting, so a possible defection cannot increase her profits, while at the same time it prolongs the length of the punishment.

Theorem 2 implies that it is not necessary to impose infinite punishment to induce the retailers' cooperation. Intuitively, a longer punishment horizon requires lower discount factors—punishment that lasts only a few periods is effective only when the retailers' discount of the future is negligible. We illustrate this with the following example.

Example 2. Suppose that $n = 3$, all three retailers face two-point demand which can achieve 0 with probability 0.5 and 10 with probability 0.5, and c_i=3.7; r_i=10; v_i=1, $i = 1,2,3; t_{ij} = 1,\ i, j = 1,2,3, i \neq j$. We have shown in Example 1 that $\delta_3^* = 0.93$ when the punishment is enforced over an infinite horizon. The value of $\delta_3^*(k)$ as a function of k is depicted in Fig. 4.1. Note that, as k increases, $\delta_3^*(k)$ approaches δ_3^*.

4.3.2 Alternative Strategies for Achieving SPNEs

Note that strategies other than the NRS described in Theorem 1 can also lead to SPNEs. One such strategy can be defined as follows: let $\mathbf{X}^{\mathrm{d}(n-1)}$ be the optimal order quantity for decentralized system with $n-1$ retailers under dual allocations.

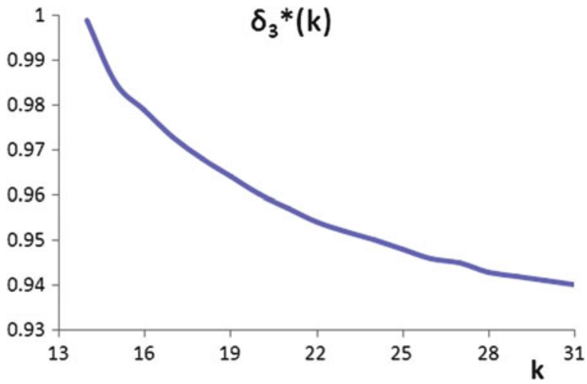


Fig. 4.1 $\delta_3^*(k)$ as a function of k

If a player, j, deviates from $(X_j^{\mathrm{d}}, \bar{H}_{jt}, \bar{E}_{jt})$ when $t = \bar{t}$, the remaining players follow strategy $(X_i^{\mathrm{d}(n-1)}, \bar{H}_{it}, \bar{E}_{it}), i \neq j, t > \bar{t}$, while retailer j adopts $(X_j^1, 0, 0), t > \bar{t}$. If a cooperating player, say l, deviates after a defection has already occurred, the punishment restarts and retailer l is excluded from future inventory sharing. Unlike the previous case (described in Theorem 1), the payoffs for the defecting player and for the cooperating players differ during the punishment period, and we need to consider them separately while checking if conditions for a SPNE are satisfied. When the discount factor is large enough, it can be shown that this strategy defines a SPNE, and that it leaves cooperating retailers with a larger payoff (during the punishment phase) than the strategy described in Theorem 1. However, observe that when the threat of punishment works, it is never actually carried out.

4.3.3 Decreasing/Increasing Costs and Prices

We would also like to mention that our model can be applied to some situations in which the costs and prices change in different periods. Let superscript t denote the values of costs/prices in period t, and suppose that $r_i^{t+1} = \rho r_i^t, v_i^{t+1} = \rho v_i^t, c_i^{t+1} = \rho c_i^t, t_{ij}^{t+1} = \rho t_{ij}^t$, for some $\rho > 0$. If $\rho < 1$, the parameters decrease with time, and our results hold if we replace δ with $\tilde{\delta} = \rho\delta$. If $\rho > 1$, the parameters increase over time, and our results will hold whenever $\tilde{\delta} < 1$, that is, when $1 < \rho < \delta^{-1}$.

4.4 Asymptotic Behavior for Large *n*

In this section, we consider the optimal retailers' ordering quantity and discount factors for large values of n. All proofs are given in the Appendix.

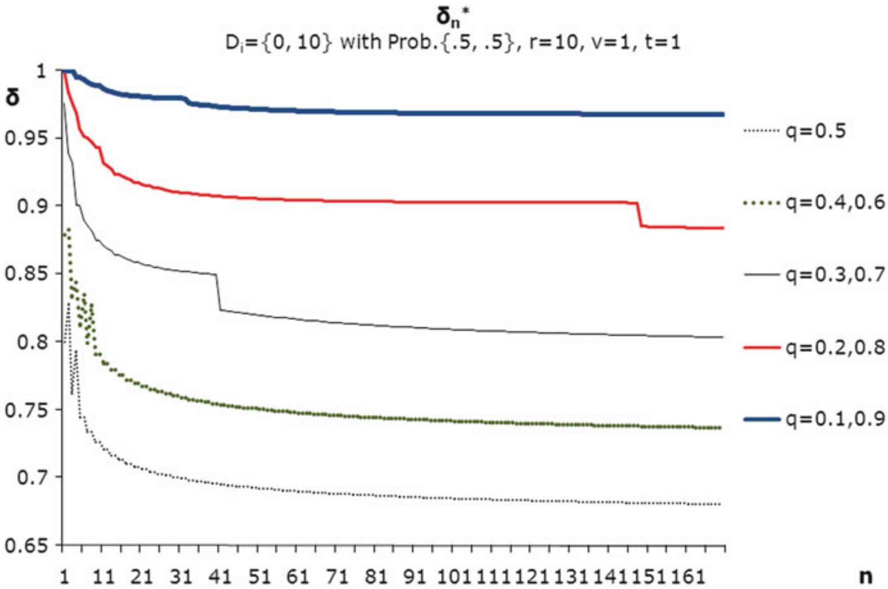


Fig. 4.2 δ_n^* for different values of the critical fractile $q = \frac{r-c}{r-v}$

We say that the retailers are *symmetric* if they face the same demand distribution F_i, cost c_i, retailer price r_i, and salvage value v_i, along with equal transportation costs in both directions, $t_{ij} = t_{ji}$. In this part of the analysis we focus on symmetric retailers, so we omit indices from notation.

Our next result provides a characterization of the lower bound for the discount factors that induces complete sharing in the NRS described in Theorem 1, δ_n^*.

Theorem 3. *In an inventory-sharing game with n symmetric retailers facing strictly increasing and independent distribution functions, there is an* $M > 0$ *such that* δ_n^* *is decreasing in n for* $n \geq \hat{n}$, *where* $\hat{n} = \min\{n \in Z : nX^{\mathrm{d}} \geq M\}$.[2]

Thus, with enough retailers participating in inventory sharing, δ_n^* is decreasing in n. Note that in many real-life situations this number can be as low as two or three. As the number of retailers increases, it is more likely for an individual retailer to benefit from inventory sharing and she is willing to participate in transshipments when she discounts her future payoffs more. We illustrate in Fig. 4.2 the behavior of δ_n^* for discrete demand that can achieve two values, 0 or 10, with equal probabilities. The two-point format of this distribution is the reason why we observe some "jumps" in the value of δ for small n. We fix r, v, and t, and change the value of c to obtain different values of the critical fractile, $q = (r-c)/(r-v)$. The values of δ_n^* are equal for "symmetric" critical fractiles

[2]If D has a finite support with upper bound $\bar{D}$, then $M = \bar{D}$.

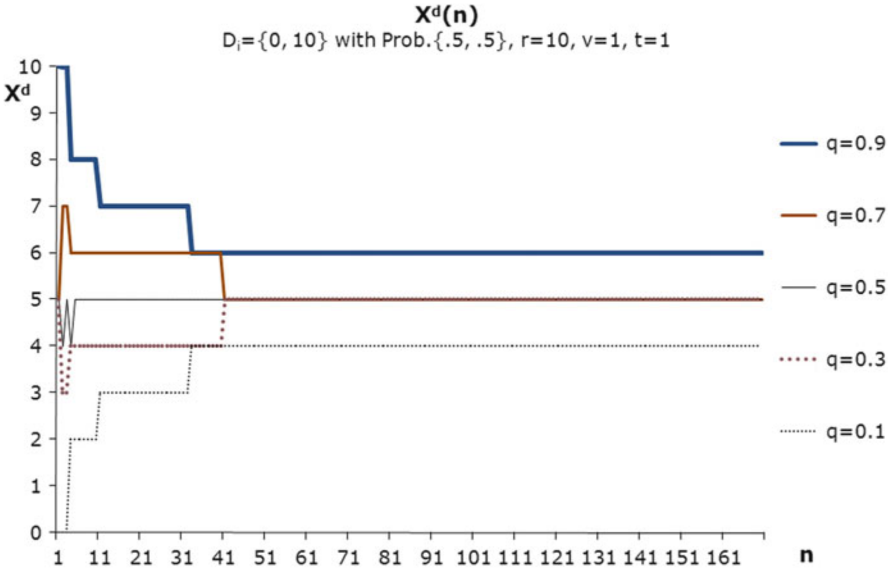


Fig. 4.3 X^{d} for different values of the critical fractile $q = \frac{r-c}{r-v}$

(q and $1-q$). As the number of retailers increases, δ_n^* shows a decreasing trend, and converges to a positive value. In addition, the discount factor that induces complete sharing increases as the critical fractile moves further from 0.5. When the critical fractile is close to 0.5, the ordering quantities at each retailer are close to the mean demand, and each retailer is more likely to benefit from transshipments. As the critical fractile moves further below (resp., above) from 0.5, each retailer orders less (resp., more), which leads to constant undersupply (resp., oversupply) and makes cooperation less useful. Therefore, cooperation is less beneficial and a larger δ is needed to incentivize the retailers.

We next characterize the retailers' ordering quantity, X^{d}.

Proposition 2. *In an inventory-sharing game with n symmetric retailers and strictly increasing distribution function $F(\cdot)$, the asymptotic behavior of the equilibrium ordering quantity can be described by*

$$\lim_{n\to\infty} X^{\mathrm{d}}(n) = \begin{cases} \mu, & \text{if } t\text{=}0 \text{ or } \frac{r-c-p}{t} \le F(\mu) \le \frac{r-c}{t}, \\ \sup\{x : F(x) < \frac{r-c}{t}\} & \text{if } F(\mu) > \frac{r-c}{t}, \\ \inf\{x : F(x) > \frac{r-c-p}{t}\} & \text{if } F(\mu) < \frac{r-c-p}{t}. \end{cases}$$

Thus, when the cost of transshipment is not too high and the margin $r-c$ is not too low, the retailers will order the mean demand value. Once again, we conduct numerical analysis with a two-point demand distribution to explore the behavior of the optimal ordering quantities and illustrate it in Fig. 4.3. One can note that in this case the optimal order quantity converges to the mean demand value, and the values corresponding to different critical fractiles are symmetric with respect to the line

$X^{\rm d} = \mu$. We note that the convergence is faster for the value of the critical fractile closer to 0.5. Due to the special nature of our demand (two-point), we may see that the optimal order quantity can exhibit some jumps initially, but eventually starts monotonic convergence toward its limit.

While in the previous case we assumed that $t = 1$ and have changed the values of c to manipulate critical fractile, we now fix the value of c and look at the impact of changes in the transshipment cost. Figure 4.4 depicts two sample cases: the graph on the left looks at the low product cost ($c = 3.7$), while the graph on the right looks at the high product cost ($c = 7.3$). In both cases, the changes in the transshipment cost determine the limiting quantity. With both low and high product cost, the limiting order quantity corresponds to the mean demand when the transshipment cost is low. However, as the transshipment cost increases, inventory sharing is less likely to occur, and the limiting order quantity moves away from the mean value—with low product cost, it moves up, and with high product cost, it moves down, which is consistent with the results from Proposition 2. When the high transshipment cost makes inventory sharing prohibitive ($t > 5$), each retailer facing high product cost (low critical fractile) orders zero, while each retailer facing low cost orders 10, which coincides with their ordering quantities without transshipments.

An immediate corollary of Proposition 2 characterizes the relationship between the retailers' optimal ordering quantities in models with and without transshipments: while the asymptotic ordering quantity may go below (resp., above) the mean demand value when the cost c becomes large (resp., small), it will never go below (resp., above) the ordering level without transshipment.

Corollary 1. *In an inventory-sharing game with n symmetric retailers and strictly increasing distribution function $F(\cdot)$, the following relationships hold when n is large:*

1. *When $t > 0$: if $F(\mu) > \frac{r-c}{t}$, then $X^1 \le X^{\rm d}(n) < \mu$; if $F(\mu) < \frac{r-c-p}{t}$, then $\mu < X^{\rm d}(n) \le X^1$.*
2. *When $t = 0$: if $F(\mu) > \frac{r-c}{r-v}$, then $X^1 \le X^{\rm d}(n) = \mu$; if $F(\mu) < \frac{r-c}{r-v}$, then $X^1 \ge X^{\rm d}(n) = \mu$.*

The results obtained so far help us in determining asymptotic behavior of δ_n^* when n is large.

Theorem 4. *In an inventory-sharing game with n symmetric retailers and strictly increasing distribution function $F(\cdot)$, $\delta_n^* \to \delta_\infty^* > 0$. More specifically, let M be as defined in Theorem 3, and let $\xi(x) = \int_0^x yf(y){\rm d}y$ and $\rho(x) = p\max\{x, M-x\}$. Then,*

$$\delta_\infty^* = \begin{cases} \frac{\rho(\mu)}{\rho(\mu)+(r-c-tF(\mu))\mu+t\xi(\mu)-(r-v)\xi(X^1)}, & \text{if } \frac{r-c-p}{t} \le F(\mu) \le \frac{r-c}{t} \text{ or } t=0; \\ \frac{\rho(X^{\rm d})}{\rho(X^{\rm d})+t\xi(X^{\rm d})-(r-v)\xi(X^1)}, & \text{if } F(\mu) > \frac{r-c}{t} \text{ and } \\ & X^{\rm d} = \sup\{x : F(x) < \frac{r-c}{t}\}; \\ \frac{\rho(X^{\rm d})}{\rho(X^{\rm d})+t(\xi(X^{\rm d})-\mu)-(r-v)(\xi(X^1)-\mu)}, & \text{if } F(\mu) < \frac{r-c-p}{t} \text{ and } \\ & X^{\rm d} = \sup\{x : F(x) > \frac{r-c-p}{t}\}. \end{cases}$$

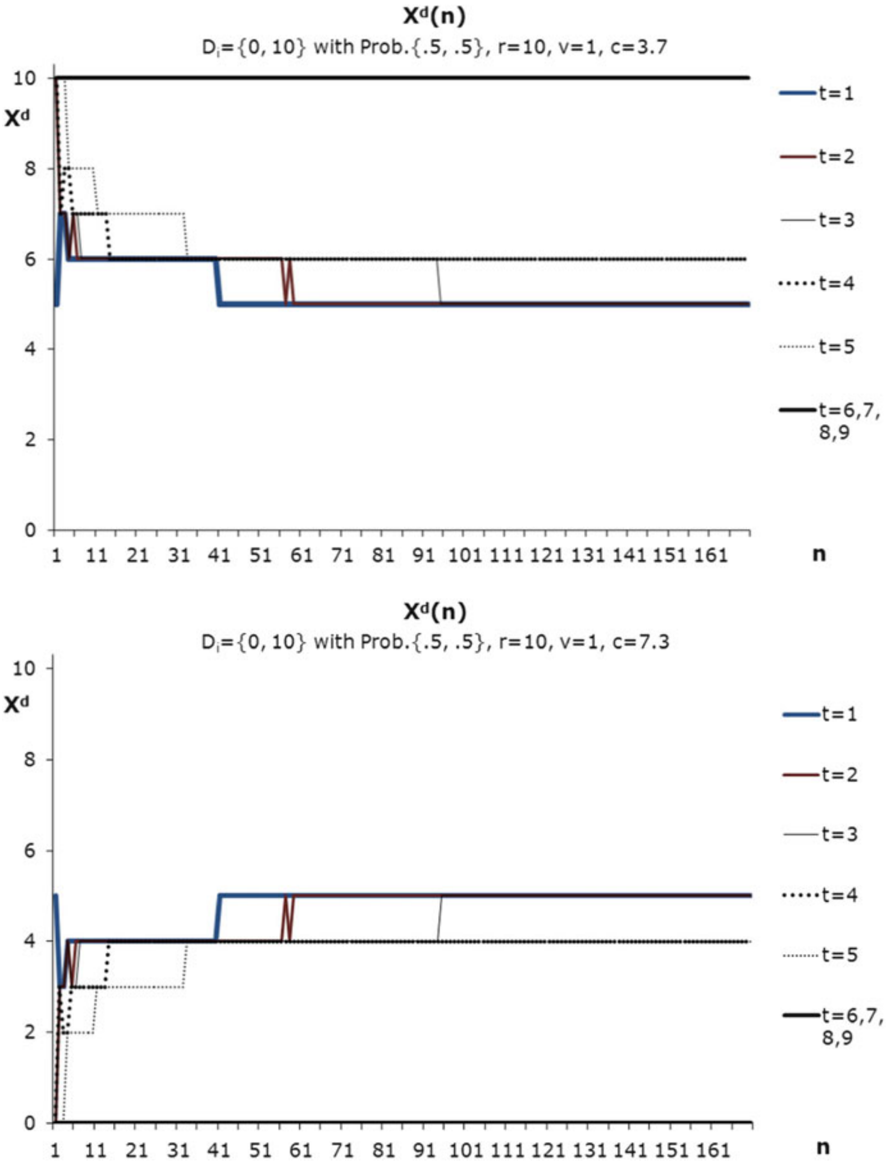


Fig. 4.4 X^{d} for different values of the transshipment cost with low and high product cost

Theorem 4 can be used to evaluate the limiting values of discount factors that induce complete sharing of residuals. An illustrative analysis is given in the following example.

Example 3. Suppose that $n \to \infty$, all retailers face demand uniformly distributed on [0,10], and $r = 10; v = 1; t = 1$. We consider different values of c, which lead to different values of the critical fractile $q = (r-c)/(r-v)$, and obtain the following results:

q	0.1	0.2	0.3	0.4	0.5	0.6	0.7	0.8	0.9
δ^*_∞	0.935	0.871	0.830	0.807	0.800	0.807	0.830	0.871	0.935

4.5 Achieving a First-Best Solution

Unfortunately, even if the retailers share all of their residuals, it is not easy to coordinate the system (except in some special cases that we discuss below) without some additional incentives, because some retailers may see a reduction in their individual profits as a result of ordering system-optimal quantities. We first discuss the cases under which a first-best outcome can be achieved without additional coordinating mechanisms, and then discuss what happens when this is not the case.

Note that even when the retailers share their entire residuals, the maximum system profit is not achieved unless the retailers order the amount optimal for the centralized model, $\mathbf{X}^n$, in each period. Thus, although the full participation and complete sharing conditions are satisfied, dual allocations may, in general, result in inefficiencies. We, therefore, start by analyzing the conditions under which decentralized stocking quantities, $\mathbf{X}^{\mathrm{d}}$, may coincide with the centralized ones, $\mathbf{X}^n$.

We first assume that the retailers are symmetric. Then, there is an equilibrium in which all retailers order the same quantity, $X_i^{\mathrm{d}} = X^{\mathrm{d}}, \forall i$, and $J_i^{\mathrm{d}}(\mathbf{X}^{\mathrm{d}}) = J^{\mathrm{d}}(\mathbf{X}^{\mathrm{d}}), \forall i$. If we consider the centralized system, there is an equilibrium in which all retailers order the same quantity, $X_i^n = X^n, \forall i$. Because the centralized model maximizes the expected profit, $J^n(\mathbf{X}^n) \geq nJ^{\mathrm{d}}(\mathbf{X}^{\mathrm{d}})$, and it is optimal for symmetric retailers to order at the first-best level, $\mathbf{X}^{\mathrm{d}} = \mathbf{X}^n$. We formalize this analysis in the following result.

Proposition 3. *If n retailers in the repeated inventory-sharing game are symmetric and $\delta > \delta_n^*$, a first-best solution can be achieved through dual allocation.*

Proposition 3 says that it is sufficient to have *symmetric* retailers to achieve a first-best outcome. This condition may be satisfied if, for instance, all retailers belong to the same organization; hence, they face the same costs/prices, and cover similar territories. However, in many realistic cases, this condition may not hold. Thus, we want to find more general conditions under which a first-best outcome can be achieved. Recall that the expected profit for retailer i is

$$J_i^{\mathrm{d}}(\mathbf{X}) = r_i \mathsf{E}[\min\{X_i, D_i\}] - c_i X_i + v_i \mathsf{E}[\bar{H}_i] + \mathsf{E}[\varphi_i^d(\mathbf{X})].$$

The total expected profit for the system of retailers is then $J^n(\mathbf{X}) = \sum_i J_i^{\mathrm{d}}(\mathbf{X})$. The optimal ordering strategy for the centralized model, $\mathbf{X}^n$, satisfies the following first-order conditions:

$$\frac{\partial J^n(\mathbf{X})}{\partial X_i} = r_i - c_i - (r_i - v_i)F_i(X_i) + \frac{\partial \mathsf{E}[\varphi_i(\mathbf{X})]}{\partial X_i} + \frac{\partial \mathsf{E}[\varphi_{-i}(\mathbf{X})]}{\partial X_i} = 0 \quad \forall i, \tag{4.2}$$

while the optimal order of an individual retailer in the decentralized system, X_i^{d}, satisfies

$$\frac{\partial J_i^{\mathrm{d}}(\mathbf{X})}{\partial X_i} = r_i - c_i - (r_i - v_i)F_i(X_i) + \frac{\partial \mathsf{E}[\varphi_i(\mathbf{X})]}{\partial X_i} = 0 \qquad \forall i. \tag{4.3}$$

Equations (4.2) and (4.3) give us a sufficient and necessary condition for a retailer in the decentralized system with an arbitrary number of retailers to order a system-optimal quantity.

Proposition 4. *If the expected total profit for the system of retailers, $J^n(\mathbf{X})$, is unimodal in $\mathbf{X}$, the sufficient and necessary condition for achieving a first-best solution is*

$$\frac{\partial \mathsf{E}[\varphi_{-i}(X_i, X_{-i}^n)]}{\partial X_i} = 0 \qquad \forall i. \tag{4.4}$$

For example, when $n = 3$, one can evaluate that the retailers with $D_i \sim U[0, 100]$; $i = 1,2,3$; $p_{12} = p_{23} = p_{31} = 6$; and $p_{21} = p_{32} = p_{13} = 8$ satisfy the above condition, and a first-best outcome can be achieved. However, through various numerical experiments we were able to observe that even small differences among parameters of different retailers may prevent us from coordinating the system. One of our analytical results is given in the following proposition.

Proposition 5. *If n retailers face i.i.d. demand distributions and differ only in their material costs (that is, $r_i = r_j = r, v_i = v_j = v, t_{ij} = t_{ji} = t$ for $i, j \in \{1, \ldots, n\}$), a first-best outcome cannot be achieved.*

We conducted a numerical analysis to study what is the impact of retailers' diversity on efficiency losses; as in Proposition 5, we assume that the retailers differ only in their cost, and study the impact of the mean and standard deviation of material cost, of the number of retailers, of the retail price, and of the salvage value. Although the system cannot be coordinated, we observe that the efficiency losses are rather small, even with a very few retailers. Some of our results are depicted in Fig. 4.5.

Our analysis indicates that, as expected, the efficiency improves as the standard deviation of cost decreases, and as the number of retailers increases. Additional simulations, in which we fix either the mean value of the cost, c, or the salvage value, v, while we vary the other parameter, indicate that the efficiency also improves with the increase of the critical fractile, which can be partially observed in Fig. 4.5. On one hand, as the decrease of the mean product cost, c, translates into larger profit

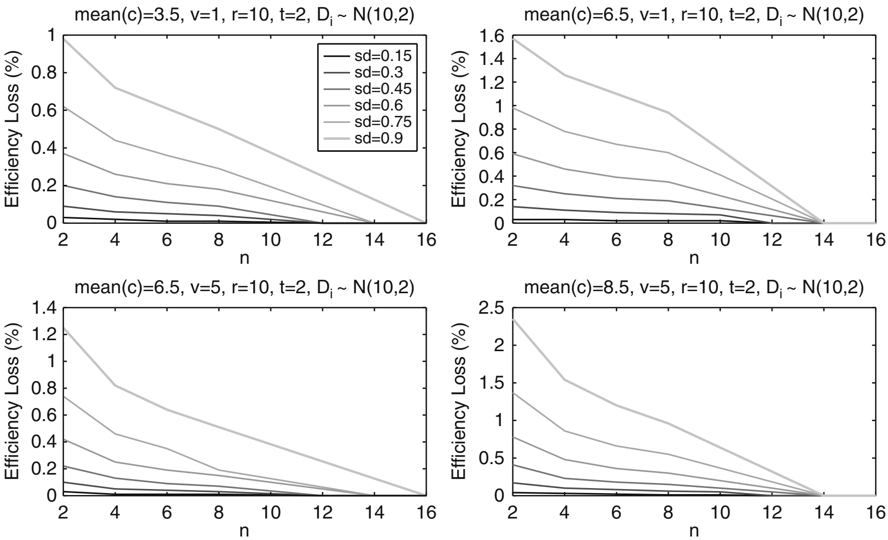


Fig. 4.5 Efficiency losses for different levels of differentiation among retailers

margin, benefits from transshipments are increased; on the other hand, the increase in the salvage value, v, hedges off the risk in demand uncertainty. In either case, the retailers' decisions become closer to those of the centralized system.

4.6 A Contractual Mechanism that Induces a First-Best Solution

In Sect. 4.5, we have shown that a first-best solution can be achieved if condition (4.4) is satisfied. However, when this condition is not satisfied, the retailers' individually optimal decisions may lead to significant efficiency losses. Achieving a first-best solution in a decentralized system may not be possible in many realistic situations without the use of some additional enforcing mechanisms.[3]

In what follows, we assume that the discount factors satisfy $\delta_i > \delta_n^*$ (hence, complete sharing is achieved), and develop a contract that leads to system-optimal order quantities without any additional constraints. Although the total system profit increases if the retailers order a first-best solution, the profit of some retailers may decrease so that they need to be induced to cooperate by some type of side payments.

[3] Note that in our repeated-game setting we were able to achieve $\mathbf{X}^{\mathrm{d}}$ as a SPNE, by utilizing the fact that $J_i^{\mathrm{d}}(X_i^D) \geq J_i^1(X_i^1)$. Unfortunately, $J_i^{\mathrm{d}}(X_i^C)$ can be greater or smaller than $J_i^1(X_i^1)$, hence a first-best ordering quantity cannot, in general, be obtained as a SPNE.

In addition, in order to prevent those retailers from defection in the future, deviations from the contract should be penalized. Thus, our contract consists of the following parts:

1. Retailer i's ordering strategy, X_{it}, and her residual-sharing amount, E_{it}, H_{it}, in every period. When a retailer orders inventory and shares residuals as prescribed by the contract, she is included in cooperation (inventory sharing) in the next period. If she breaks the contract and orders a different quantity or shares a different amount in period t, she is excluded from cooperation in all subsequent periods, $t+1, t+2, \ldots$
2. Discretionary transfer payments at the end of each period. A retailer, i, who breaks the contract in period t makes a positive payment, d_{it}. This value is distributed among the retailers who have followed the contract, $\sum_i d_{it} = 0$.
3. Contract activation bonus, B_i, upon signing the contract. This one-time bonus can be positive or negative, with $\sum_i B_i = 0$. The retailers who benefit from cooperation are those that may be required to have a negative activation bonus in order to induce participation of retailers who would individually prefer not to order a first-best quantity.

We will refer to this contract as the *eviction contract* because the most severe punishment for a defecting retailer is her eviction from the inventory-sharing system. Changes in the cooperative behavior of the system can be described through coalition structures, in which cooperating retailers belong to a coalition. Each time a retailer is evicted, the remaining retailers form a new inventory-sharing system and completely share residuals in this reduced system. Thus, if none of the retailers has ever defected, the system operates as the grand coalition. We assume that the retailers who are evicted do not form new inventory-sharing groups. This implies that each evicted retailer constitutes a one-member coalition. In other words, suppose that the current system is described by coalition structure $Z = \{S_1, S_2, \ldots, S_{n-k+1}\}$. Then, $|S_j| = 1$ for $n-k$ coalitions, and $|S_i| = k$ for some coalition S_i. We will use Z^k to denote a coalition structure in which exactly one coalition has k members, while the remaining $n-k$ coalitions consist of a single retailer. Thus, Z^n denotes the grand coalition, while Z^1 denotes the coalition structure with no inventory sharing. Clearly, the system-optimal stocking quantity in state Z^n is $\mathbf{X}^n$, while $\mathbf{X}^1$ maximizes the system profit under state Z^1. We denote by $\mathbf{X}^k$ the system-optimal stocking quantities for coalition structure Z^k. For an arbitrary coalition structure, Z, we denote the system-optimal order quantity by $\mathbf{X}^Z$.

In order to induce a system-optimal solution, the eviction contract requires the retailers to order system-optimal quantities and share all of their residuals. Thus, given a coalition structure, the orders placed and residuals shared by the retailers in period t, we can determine the coalition structure in period $t+1$ as follows:

$$Z_{t+1}(Z_t, \mathbf{X}_t, \mathbf{H}_t, \mathbf{E}_t | Z_t = Z^k) = \begin{cases} Z^k, & \text{if } \mathbf{X}_t = \mathbf{X}^k, \mathbf{H}_t = \bar{\mathbf{H}}_t, \mathbf{E}_t = \bar{\mathbf{E}}_t; \\ Z^{k-l}, & \text{if } (X_{it} = X_i^k, H_{it} = \bar{H}_{it}, E_{it} = \bar{E}_{it}) \text{ does} \\ & \text{not hold for } l \text{ coalition members in } Z_t. \end{cases} \tag{4.5}$$

Now, the eviction contract can be described by

$$(\mathbf{X}_t(\hat{\mathbf{h}}_{t-1}), \mathbf{H}_t(\hat{\mathbf{h}}_{t-1}), \mathbf{E}_t(\hat{\mathbf{h}}_{t-1}), \mathbf{d}_t(\hat{\mathbf{h}}_t), \mathbf{B}),$$

where $\hat{\mathbf{h}}_t$ denotes the history up to period t, $\hat{\mathbf{h}}_t = \{Z_\tau, \mathbf{X}_\tau, \mathbf{H}_\tau, \mathbf{E}_\tau\}_{\tau=1}^{t}$.

Recall that we use J_i^{d} to denote the expected profit for retailer i under dual allocations when all retailers participate in inventory sharing. We now introduce some additional notation. We denote by $J_i^Z(\mathbf{X}, \mathbf{H}, \mathbf{E})$ the expected profit for i under coalition structure Z, and by $J^Z(\mathbf{X}, \mathbf{H}, \mathbf{E})$ the expected total system profit under coalition structure Z. The following theorem describes how a first-best solution can be achieved through an eviction contract. Its proof is given in the Appendix.

Theorem 5. *Suppose that all retailers participate in inventory sharing and $J^n(\mathbf{X})$ is unimodal. Then, the eviction contract $(\mathbf{X}_t(\hat{\mathbf{h}}_{t-1}), \mathbf{H}_t(\hat{\mathbf{h}}_{t-1}), \mathbf{E}_t(\hat{\mathbf{h}}_{t-1}), \mathbf{d}_t(\hat{\mathbf{h}}_t), \mathbf{B})$ is a contract that induces a first-best solution if the retailers' ordering strategies, $\mathbf{X}_t$, are given by*

$$\mathbf{X}_t(\hat{\mathbf{h}}_{t-1} | Z_t = Z^k) = \mathbf{X}^k,$$

all coalition members share their entire residuals, the evicted members share nothing, the discretionary transfer payments are

$$d_{it}(\hat{\mathbf{h}}_t) = \begin{cases} \frac{\Delta_{it}(\hat{\mathbf{h}}_t)}{\sum_{I_t^+} \Delta_{jt}(\hat{\mathbf{h}}_t)} \times \sum_{I_t^-} (-\Delta_{jt}(\hat{\mathbf{h}}_t)) & i \in I_t^+ \\ \Delta_{it}(\hat{\mathbf{h}}_t) & i \in I_t^-, \end{cases}$$

where

$$\Delta_{it}(\hat{\mathbf{h}}_t) = \frac{1}{1-\delta_i}\left[J_i^{Z_t}(\mathbf{X}^{Z_t}) - \delta J_i^1(X_i^1)\right] - J_i^{Z_t}(\mathbf{X}_t, \mathbf{H}_t, \mathbf{E}_t),$$

$$I_t^+ = \{i : \Delta_{it}(\hat{\mathbf{h}}_t) > 0\} \text{ and } I_t^- = \{i : \Delta_{it}(\hat{\mathbf{h}}_t) \le 0\},$$

and the one-time contract activation bonus is given as

$$B_i = \begin{cases} \frac{\Lambda_i}{\sum_{K^-} \Lambda_i} \times \sum_{K^+} (-\Lambda_i) & i \in K^- \\ \Lambda_i & i \in K^+, \end{cases}$$

where

$$\Lambda_i = \frac{1}{1-\delta_i}(J_i^n(\mathbf{X}^{\mathrm{d}}) - J_i^n(\mathbf{X}^n)), \;\; K^+ = \{i : \Lambda_i > 0\} \text{ and } K^- = \{i : \Lambda_i \le 0\}.$$

Despite its seemingly complex structure, the contract is actually quite simple to implement: at the beginning of their cooperation, the retailers who strictly benefit from the contract compensate the retailers whose profit is reduced (as a result of ordering system-optimal quantities) through the activation bonus B_i. In addition, the retailers agree that in the case of any defection, all benefits should be forfeited and allocated among the retailers who suffer a loss after such an action.[4] Thus, there is no incentive for any retailer to defect from the strategy which prescribes ordering system-optimal quantity, sharing entire residuals, and receiving dual allocations. The transfer payment is zero as long as the retailers follow the contract—it serves as a threat that prevents them from defection.[5] One could, alternatively, develop a contract in which retailers whose profit decreases after ordering system-optimal quantity receive compensations for their losses at the end of every period. This type of contract would not require activation bonuses, but may lead to more complex implementation, as the payments need to be calculated and exchanged at the end of every period (in our contract, this happens only if there was a defection in a given period).

Note that the eviction contract works not only for dual allocations, but also for any other allocation rule that induces full participation and complete residual sharing, but not a first-best inventory decision. This can be easily confirmed by observing that the proof does not depend upon any pre-specified allocation rules.

4.7 Concluding Remarks

In this work, we study a repeated inventory-sharing game with n retailers in which the retailers distribute the profit from transshipments according to the dual allocations. Each retailer faces stochastic demand and salvages all unsold inventory at the end of each period. Using the standard tools from the theory of repeated games, we show that the use of NRS induces complete sharing in an SPNE of an infinitely repeated game (providing that the discount factor of future payoffs is large enough), while the retailers always withhold residuals if the game is repeated a finite number of times. We also show that complete sharing can be an SPNE even if the punishment is not executed over an infinite horizon but instead lasts only for a finite number of periods. Clearly, shorter punishment periods require larger discount factors, and a punishment that lasts only a few periods will induce complete sharing only with the retailers whose discounting of the future periods is very small. In addition, we provide some analytical results for the asymptotic behavior of the retailers' ordering quantities and the lower bounds on discount factors that induce complete sharing for large number of players.

---

[4]The amount of transfer payments $d_i \leq 0$ (realized when a player benefits from a defection) removes from a retailer all possible gains from that defection. $\Delta_{it} > 0$ (which leads to $d_i > 0$) implies that a retailer observes a loss as a result of someone's defection (and is, therefore, compensated from payments of those who benefit); this retailer receives a fraction of total transfer payments proportional to her loss as compared to the total losses observed by the system.

[5]In the whole contract lifetime, the discretionary transfer payment happens at most $n-1$ times, as the number of inventory-sharing retailers is reduced from n to 1.

While there can be a significant difference in optimal profits generated by decentralized retailers and those generated in a centralized system, a decentralized model will result in a system-optimal outcome if the retailers are symmetric. As this condition may not be satisfied in many cases, we derive another condition, (4.4), that leads to a first-best outcome. When this condition is not satisfied, we develop a contract that induces the retailers to order a first-best quantity whenever the complete sharing condition holds.

We note that our model assumes that all leftover inventory is salvaged at the end of each period. The reason for this is twofold. On one hand, because we were mainly interested in studying the impact of repeated interactions on the retailers' sharing decisions in the second stage, a more complex model in which the retailers are allowed to carry inventory from one period to another would lead to a more complicated model that is beyond the scope of this work. On another hand, such situations do occur in industries where products have short life cycles, long lead times, and unpredictable demands, like apparel, Christmas toys, and high-tech electronic components. Retailers in these industries are often open to inventory-sharing agreements with others.

Our inventory-sharing model may require a neutral third party for its implementation—monitoring of residuals, making effective transshipment decisions, and allocation of profits among the members. While this is easily realized within a trade association or when the retailers belong to a larger organization, it might be more difficult to execute when the retailers are independent. It is, therefore, interesting to observe emergence of companies such as iSuppli Corp., which act as neutral intermediaries among independent entities and, at the same time, improve the market's efficiency.

When dual allocations are used in one-shot setting, the retailers withhold their residuals, and our aim was to study if this property persists when the retailers interact repeatedly. Note, however, that many of our results can be extended to alternative allocation rules (though some extensions may require certain modifications in proofs and results).

Appendix

Proof of Theorem 3. In order to prove this theorem, we first introduce the following notation: let $F^m(y) = P\{\sum_{i=1}^m D_i \le y\}$, $\hat{F}^m(y) = P\{\frac{1}{m}\sum_{i=1}^m D_i \le y\}$, and $\mathsf{E}[D_i] = \mu$. Note that $F^m(y) = \hat{F}^m(\frac{y}{m})$. We will also need the following lemmas.

Lemma 1. *In an inventory-sharing game with symmetric retailers facing strictly increasing and independent distribution functions, a retailer defecting from strategy* $(X^{\mathrm{d}}, \bar{H}_i, \bar{E}_i)$ *maximizes her benefit from defection if she orders* X^{d}.

Proof of Lemma 1. If we have n symmetric retailers, the dual price of retailer i's residual will be either 0 or p, depending on the amount she is sharing with the others.

For example, if $\sum_{j\neq i}(\bar{E}_j - \bar{H}_j) = k > 0$, the retailers other than i need k additional units of products. Then, retailer i will receive p per unit if $0 < \bar{H}_i < k$, while she will get nothing otherwise. More formally, retailer i's total expected profit when she orders X_i and other retailers order $\mathbf{X}^{\mathrm{d}}_{-i}$ is given by

$$\begin{aligned}
J_i^{\mathrm{d}}(X_i|\mathbf{X}^{\mathrm{d}}_{-i}) = {} & r\mathsf{E}[\min\{X_i, D_i\}] + v\mathsf{E}[H_i] - cX_i \\
& + p\int_0^{\infty} f^{n-1}\left((n-1)X^{\mathrm{d}} + k\right)\int_{X_i-k}^{X_i}(X_i - u)f(u)\mathrm{d}u\mathrm{d}k \\
& + p\int_0^{\infty} f^{n-1}\left((n-1)X^{\mathrm{d}} - k\right)\int_{X_i}^{X_i+k}(u - X_i)f(u)\mathrm{d}u\mathrm{d}k,
\end{aligned}$$

where $f^{n-1}((n-1)X^{\mathrm{d}} + y)$ is the probability density when the residual demand (resp., inventory) for the remaining $(n-1)$ retailers is $y > 0$ (resp., $(-y) > 0$), and its first derivative is given by

$$\begin{aligned}
(J_i^{\mathrm{d}})'(X_i|\mathbf{X}^{\mathrm{d}}_{-i}) = {} & r - c - (r-v)F(X_i) \\
& + p\int_0^{\infty}[F(X_i) - F(X_i - k)]f^{n-1}\left((n-1)X^{\mathrm{d}} + k\right)\mathrm{d}k \\
& - p\int_0^{\infty}[F(X_i + k) - F(X_i)]f^{n-1}\left((n-1)X^{\mathrm{d}} - k\right)\mathrm{d}k \\
& - p\int_0^{\infty} k\Big[f(X_i - k)f^{n-1}\left((n-1)X^{\mathrm{d}} + k\right) \\
& - f(X_i + k)f^{n-1}\left((n-1)X^{\mathrm{d}} - k\right)\Big]\mathrm{d}k. \qquad (4.6)
\end{aligned}$$

Retailer i can increase her profit if she deviates whenever her dual price is zero. In other words, she maximizes her profit if she withholds part of her residual inventory/demand to make it lower than the total residual demand/inventory from other retailers. Under this kind of strategy, her total expected profit will be increased to

$$\begin{aligned}
J_i^{\mathrm{def}}(X_i|\mathbf{X}^{\mathrm{d}}_{-i}) = {} & r\mathsf{E}[\min\{X_i, D_i\}] + v\mathsf{E}[H_i] - cX_i \\
& + p\int_0^{\infty} f^{n-1}\left((n-1)X^{\mathrm{d}} + k\right)\int_{X_i-k}^{X_i}(X_i - u)f(u)\mathrm{d}u\mathrm{d}k \\
& + p\int_0^{\infty} f^{n-1}\left((n-1)X^{\mathrm{d}} - k\right)\int_{X_i}^{X_i+k}(u - X_i)f(u)\mathrm{d}u\mathrm{d}k \\
& + p\int_0^{\infty} k f^{n-1}\left((n-1)X^{\mathrm{d}} + k\right)F(X_i - k)\mathrm{d}k \\
& + p\int_0^{\infty} k f^{n-1}\left((n-1)X^{\mathrm{d}} - k\right)[1 - F(X_i + k)]\mathrm{d}k,
\end{aligned}$$

and its derivatives are

$$(J_i^{\text{def}})'(X_i|\mathbf{X}_{-i}^{\text{d}}) = r - c - (r-v)F(X_i) + p\int_0^{\infty} [F(X_i) - F(X_i - k)]\, f^{n-1}\left((n-1)X^{\text{d}} + k\right) \text{d}k - p\int_0^{\infty} [F(X_i + k) - F(X_i)]\, f^{n-1}\left((n-1)X^{\text{d}} - k\right) \text{d}k, \quad (4.7)$$

$$(J_i^{\text{def}})''(X_i|\mathbf{X}_{-i}^{\text{d}}) = -t f(X_i) - p\int_0^{\infty} \Big[f(X_i - k) f^{n-1}\left((n-1)X^{\text{d}} + k\right) + f(X_i + k) f^{n-1}\left((n-1)X^{\text{d}} - k\right)\Big] \text{d}k < 0. \quad (4.8)$$

Because all demands follow an identical distribution, it follows from (4.6) and (4.7) that

$$\begin{aligned}
&[(J_i^{\text{def}})' - (J_i^{\text{d}})'](X_i|\mathbf{X}_{-i}^{\text{d}}) \\
&= p\int_0^{\infty} k\Big[f(X_i - k) f^{n-1}\left((n-1)X^{\text{d}} + k\right) - f(X_i + k) f^{n-1}\left((n-1)X^{\text{d}} - k\right)\Big] \text{d}k \\
&= E\left[X_i - D_i |\sum_{m=1}^{n} D_m = (n-1)X^{\text{d}} + X_i\right] \\
&= \frac{n-1}{n}\left(X_i - X^{\text{d}}\right).
\end{aligned}$$

Recall that $X^{\text{d}} = \arg\max J_i^{\text{d}}(X_i|\mathbf{X}_{-i}^{\text{d}})$, and consequently $(J_i^{\text{d}})'(X^{\text{d}}|\mathbf{X}_{-i}^{\text{d}})) = 0$. This implies

$$\begin{aligned}
(J_i^{\text{def}})'(X^{\text{d}}|\mathbf{X}_{-i}^{\text{d}}) &= (J_i^{\text{d}})'(X^{\text{d}}|\mathbf{X}_{-i}^{\text{d}}) + [(J_i^{\text{def}})'(X^{\text{d}}|\mathbf{X}_{-i}^{\text{d}}) - (J_i^{\text{d}})'(X^{\text{d}}|\mathbf{X}_{-i}^{\text{d}})] \\
&= 0 + \frac{n-1}{n}(X^{\text{d}} - X^{\text{d}}) = 0.
\end{aligned}$$

Since $J_i^{\text{def}}(X_i|\mathbf{X}_{-i}^{\text{d}})$ is a concave function, the optimal ordering decision when player i defects, X_i^{def}, should satisfy $(J_i^{\text{def}})'(X_i^{\text{def}}|\mathbf{X}_{-i}^{\text{d}})) = 0$. Thus, $X_i^{\text{def}} = X^{\text{d}}$, and a retailer contemplating a defection maximizes her profit if she orders at the decentralized optimal level. □

Lemma 2. *In an inventory-sharing game with n symmetric retailers and strictly increasing demand distribution function, the expected profit for each retailer,* $J^{\text{d}}\left(X^{\text{d}}(n), n\right)$*, is increasing in n, where* $X^{\text{d}}(n)$ *is the NE ordering decision for each retailer in the decentralized system.*

Proof of Lemma 2. Consider a game with $n+1$ symmetric retailers, and let $\mathscr{S}$ be any n-members subset of these retailers. In terms of cooperative game theory, the value of the coalition $\mathscr{S}$ corresponds to the profit generated by its members; because the retailers are symmetric, it can be written as $V^*_{\mathscr{S}} = nJ^{\mathrm{d}}(X,n)$, where $J^{\mathrm{d}}(X,n)$ denotes the expected profit generated by an arbitrary retailer in a game with n symmetric retailers under dual allocations. However, in an $(n+1)$-retailer game with dual allocations, each retailer will receive a payoff $J^{\mathrm{d}}(X,n+1)$. Because dual allocations belong to the core, we must have $nJ^{\mathrm{d}}(X,n+1) > V^*_{\mathscr{S}} = nJ^{\mathrm{d}}(X,n)$. It is then straightforward that $J^{\mathrm{d}}\left(X^{\mathrm{d}}(n+1),n+1\right) \geq J^{\mathrm{d}}\left(X^{\mathrm{d}}(n),n+1\right) \geq J^{\mathrm{d}}\left(X^{\mathrm{d}}(n),n\right)$. □

We can now prove the theorem. Consider the model with n symmetric retailers and suppose that there were no prior defections. That is, each retailer orders X^{d} and shares her entire residuals. Recall that we have shown in Lemma 1 that defecting retailers maximize their profit if they order X^{d} and deviate in the amount they share with others. Under demand realization $\mathbf{D}$, let $\bar{P}_i^{\mathrm{def}}(\mathbf{X}^{\mathrm{d}},\mathbf{D},n)$ denote the highest payoff that retailer i can generate if she defects in a game with n players, while the other retailers cooperate, and recall that $P_i^d(\mathbf{X}^{\mathrm{d}},\mathbf{D},n)$ is her profit in the current period if she shares all of her residuals. After defection, she will receive $J_i(X_1)$ in all subsequent periods. Thus, a possible deviation by player i is deterred if her discount factor satisfies

$$\bar{P}_i^{\mathrm{def}}(\mathbf{X}^{\mathrm{d}},\mathbf{D},n) + \frac{\delta}{1-\delta}J_i(X_1) < \frac{\delta}{1-\delta}J_i^{\mathrm{d}}(\mathbf{X}^{\mathrm{d}},n) + P_i^d(\mathbf{X}^{\mathrm{d}},\mathbf{D},n), \forall \mathbf{D}, \tag{4.9}$$

where $J_i^{\mathrm{d}}(\mathbf{X}^{\mathrm{d}},n)$ denotes the payoff that retailer i receives when n retailers use dual allocations, order $\mathbf{X}^{\mathrm{d}}$, and share their entire residuals. It is easy to verify that (4.9) holds whenever

$$\delta > \delta_{i,n} = \frac{1}{1+\frac{J_i^{\mathrm{d}}(\mathbf{X}^{\mathrm{d}},n)-J_i(X_1)}{\sup_{\mathbf{D}}\{\bar{P}_i^{\mathrm{def}}(\mathbf{X}^{\mathrm{d}},\mathbf{D},n)-P_i^d(\mathbf{X}^{\mathrm{d}},\mathbf{D},n)\}}}. \tag{4.10}$$

Note that the upper bound of the extra profit that one can get out of deviation, $\sup_{\mathbf{D}}\{\bar{P}_i^{\mathrm{def}}(\mathbf{X}^{\mathrm{d}},\mathbf{D},n) - P_i^d(\mathbf{X}^{\mathrm{d}},\mathbf{D},n)\}$, can be obtained by comparing two cases: (1) the extra profit generated when $D_i = 0$ and the total residual demand of the remaining retailers is slightly below X^{d}; and (2) the extra profit generated when $D_{-i} = 0$ and D_i is slightly above nX^{d}. In the first case, this profit is pX^{d}; in the second case, this profit would be $p(n-1)X^{\mathrm{d}}$, assuming that demand can achieve values above nX^{d}. However, note that in most real-life situations there is an $M > 0$ such that $P(D_i > M)$ is negligible (if demand distribution has a finite support with upper bound $\bar{D}$, then $M = \bar{D}$), and the maximum benefit from defection is $p(M - X^{\mathrm{d}})$. Let us denote $\hat{n} = \min\{n : nX^{\mathrm{d}} \geq M\}$. Then, whenever $n \geq \hat{n}$, it implies that $\sup_{\mathbf{D}}\{\bar{P}_i^{\text{def}}(\mathbf{X}^{\text{d}},\mathbf{D},n) - P_i^d(\mathbf{X}^{\text{d}},\mathbf{D},n)\} = \max\{pX^{\text{d}}, p(M-X^{\text{d}})\}$, and (4.10) corresponds to

$$\delta > \delta_{i,n} = \frac{p\max\{X^{\text{d}}, M-X^{\text{d}}\}}{p\max\{X^{\text{d}}, M-X^{\text{d}}\} + J_i^{\text{d}}(\mathbf{X}^{\text{d}},n) - J_i(X_1)}.$$

Because the players are symmetric, let $\delta_n = \delta_{i,n}$. Since $J_i(X_1)$ does not depend on n and we showed in Lemma 2 that $J_i^{\text{d}}(\mathbf{X}^{\text{d}},n)$ increases with n, δ_n is decreasing in n. Finally, let $\delta_n^* = \delta_n$. □

Proof of Proposition 2. When each retailer orders X^{d}, the total expected profit for each of them can be determined by

$$\begin{aligned} J(\mathbf{X}^{\text{d}}) &= r\mathsf{E}[\min\{X^{\text{d}}, D\}] + v\mathsf{E}[H] - cX^{\text{d}} \\ &\quad + p\int_0^\infty kf(X^{\text{d}}-k)\left[1-\hat{F}^{n-1}\left(X^{\text{d}}+\frac{k}{n-1}\right)\right]\mathrm{d}k \\ &\quad + p\int_0^\infty kf(X^{\text{d}}+k)\hat{F}^{n-1}\left(X^{\text{d}}-\frac{k}{n-1}\right)\mathrm{d}k \\ &= (r-c)X^{\text{d}} - (r-v)\left[X^{\text{d}}F(X^{\text{d}}) - \int_0^{X^{\text{d}}} yf(y)\mathrm{d}y\right] \\ &\quad + p\int_0^\infty kf(X^{\text{d}}-k)\left[1-\hat{F}^{n-1}\left(X^{\text{d}}+\frac{k}{n-1}\right)\right]\mathrm{d}k \\ &\quad + p\int_0^\infty kf(X^{\text{d}}+k)\hat{F}^{n-1}\left(X^{\text{d}}-\frac{k}{n-1}\right)\mathrm{d}k. \end{aligned}$$

If we let $\sigma^2 = Var[D_i]$, then by the central limit theorem (CLT) we have

$$\lim_{m\to\infty}\frac{1}{m}\sum_{i=1}^{m} D_i \sim N\left(\mu, \frac{\sigma^2}{m}\right).$$

Suppose first that $X^{\text{d}} > \mu$. Then, we have $\lim_{n\to\infty}[1-\hat{F}^{n-1}(X^{\text{d}}+\frac{k}{n-1})] = 0$ and $\lim_{n\to\infty}\hat{F}^{n-1}(X^{\text{d}}-\frac{k}{n-1}) = 1$; hence, the derivative of $J(\cdot|\mathbf{X}_{-i}^{\text{d}})$ evaluated at X^{d} becomes

$$J'(X^{\text{d}}|\mathbf{X}_{-i}^{\text{d}}) = r-c-(r-v)F(X^{\text{d}}) - p + pF(X^{\text{d}}) = -(c-v) + t[1-F(X^{\text{d}})],$$

which is a decreasing function of X^{d}. Thus, if $t=0$ or $F(\mu) \geq 1-\frac{c-v}{t} = \frac{r-c-p}{t}$, then $J'(X^{\text{d}}|\mathbf{X}_{-i}^{\text{d}}) \leq 0$ for any $X^{\text{d}} \in (\mu,\infty)$, and the retailer maximizes her profit by choosing $X^{\text{d}} \to \mu^+$. Otherwise, $X^{\text{d}} = \inf\{x : F(x) > \frac{r-c-p}{t}\}$ is an optimal solution within (μ,∞).

If $X^{\mathrm{d}} < \mu$, $\lim_{n\to\infty}[1-\hat{F}^{n-1}(X^{\mathrm{d}}+\frac{k}{n-1})] = 1$ and $\lim_{n\to\infty}\hat{F}^{n-1}(X^{\mathrm{d}}-\frac{k}{n-1}) = 0$. The derivative of $J(\cdot|\mathbf{X}^{\mathrm{d}}_{-i})$ evaluated at X^{d} becomes

$$J^{'}(X^{\mathrm{d}}|\mathbf{X}^{\mathrm{d}}_{-i}) = r-c-(r-v)F(X^{\mathrm{d}})+pF(X^{\mathrm{d}}) = (r-c)-tF(X^{\mathrm{d}}),$$

which is again a decreasing function of X^{d}. In this case, if $F(\mu) \le \frac{r-c}{t}$ or $t = 0$, then $J^{'}(X^{\mathrm{d}}|\mathbf{X}^{\mathrm{d}}_{-i}) \ge 0$ for any $X^{\mathrm{d}} \in (\infty,\mu)$, and the retailer maximizes her profit by choosing $X^{\mathrm{d}} \to \mu^{-}$. Otherwise, $X^{\mathrm{d}} = \sup\{x : F(x) < \frac{r-c}{t}\}$ is an optimal solution within $(-\infty,\mu)$.

From the above, we can conclude that whenever $F(\mu) \in [\frac{r-c-p}{t}, \frac{r-c}{t}]$ or $t = 0$, the retailer should select $X^{\mathrm{d}} \to \mu$. Otherwise, because $\frac{r-c-p}{t} \le \frac{r-c}{t}$, any local optimum is also a global optimum whenever $F(\mu) \notin [\frac{r-c-p}{t}, \frac{r-c}{t}]$. □

Proof of Corollary 1. Suppose first that $t > 0$. If $F(\mu) > \frac{r-c}{t}$, it follows from Proposition 2 that $\lim_{n\to\infty} X^{\mathrm{d}}(n) = \sup\{x : F(x) < \frac{r-c}{t}\}$. This implies that $F(X^{\mathrm{d}}) \le \frac{r-c}{t} < F(\mu)$, hence $X^{\mathrm{d}} < \mu$. On the other hand, when there is no cooperation among the retailers, the optimal ordering level X^1 can be determined by the newsvendor model, $F(X^1) = \frac{r-c}{r-v}$. Recall that we assume $p = r-v-t \ge 0$, which implies $r-v \ge t$, therefore $F(X^1) \le F(X^{\mathrm{d}})$, and $X^1 \le X^{\mathrm{d}}$.

If, on the other hand, $F(\mu) < \frac{r-c-p}{t}$, then $\lim_{n\to\infty} X^{\mathrm{d}}(n) = \inf\{x : F(x) > \frac{r-c-p}{t}\}$. This implies that $F(\mu) < \frac{r-c-p}{t} \le F(X^{\mathrm{d}})$, hence $\mu < X^{\mathrm{d}}$. Consequently, $F(X^1) = \frac{r-c}{r-v} \ge \frac{r-c-p}{r-v-p} = \frac{r-c-p}{t} = F(X^{\mathrm{d}})$, so $X^1 \ge X^{\mathrm{d}}$.

When $t = 0$, each retailer orders the expected demand value, and the result is straightforward. □

Proof of Theorem 4. Recall that the lower bound of δ_n satisfies

$$\begin{aligned}\delta_n^* &= \frac{p\max\{X^{\mathrm{d}}, M-X^{\mathrm{d}}\}}{p\max\{X^{\mathrm{d}}, M-X^{\mathrm{d}}\}+J_i^{\mathrm{d}}(\mathbf{X}^{\mathrm{d}},n)-J_i(X_1)} \\ &= \frac{\rho(X^{\mathrm{d}})}{\rho(X^{\mathrm{d}})+J_i^{\mathrm{d}}(\mathbf{X}^{\mathrm{d}},n)-J_i(X_1)}\forall i. \end{aligned} \tag{4.11}$$

In addition, in the model without cooperation, each retailer's profit is maximized at $X^1 = F^{-1}\left(\frac{r-c}{r-v}\right)$, and equals

$$J^1(X^1) = (r-v)\int_0^{X^1} yf(y)\mathrm{d}y = (r-v)\xi(X^1). \tag{4.12}$$

If $X^{\mathrm{d}} = \mu$, it follows from the CLT that

$$\lim_{n\to\infty} 1-\hat{F}^{n-1}\left(X^{\mathrm{d}}+\frac{k}{n-1}\right) = \lim_{n\to\infty}\hat{F}^{n-1}\left(X^{\mathrm{d}}-\frac{k}{n-1}\right) = \frac{1}{2},$$

which implies

$$
\begin{aligned}
J_i^{\mathrm{d}}(\mathbf{X}^{\mathrm{d}},n) &= (r-c)X^{\mathrm{d}} - (r-v)\left[X^{\mathrm{d}}F(X^{\mathrm{d}}) - \int_0^{X^{\mathrm{d}}} yf(y)\mathrm{d}y\right] \\
&\quad + p\int_0^{\infty} kf(X^{\mathrm{d}}-k)\left[1-\hat{F}^{n-1}\left(X^{\mathrm{d}}+\frac{k}{n-1}\right)\right]\mathrm{d}k \\
&\quad + p\int_0^{\infty} kf(X^{\mathrm{d}}+k)\hat{F}^{n-1}\left(X^{\mathrm{d}}-\frac{k}{n-1}\right)\mathrm{d}k \\
&= (r-c)\mu - (r-v)\left[\mu F(\mu) - \int_0^{\mu} yf(y)\mathrm{d}y\right] \\
&\quad + \frac{p}{2}\left[\int_0^{\infty} kf(\mu-k)\mathrm{d}k + \int_0^{\infty} kf(\mu+k)\mathrm{d}k\right] \\
&= [r-c-tF(\mu)]\mu + t\int_0^{\mu} yf(y)\mathrm{d}y \\
&= [r-c-tF(\mu)]\mu + t\xi(\mu).
\end{aligned} \tag{4.13}
$$

By substituting (4.12) and (4.13) into (4.11), we obtain

$$
\delta_{\infty}^{*} = \frac{\rho(\mu)}{\rho(\mu) + [r-c-tF(\mu)]\mu + t\xi(\mu) - (r-v)\xi(X^1)}.
$$

If $X^{\mathrm{d}} = \sup\{x : F(x) < \frac{r-c}{t}\} < \mu$, we have $\lim_{n\to\infty} 1 - \hat{F}^{n-1}(X^{\mathrm{d}} + \frac{k}{n-1}) = 1$ and $\lim_{n\to\infty} \hat{F}^{n-1}(X^{\mathrm{d}} - \frac{k}{n-1}) = 0$, hence

$$
\begin{aligned}
J_i^{\mathrm{d}}(\mathbf{X}^{\mathrm{d}},n) &= (r-c)X^{\mathrm{d}} - (r-v)\left[X^{\mathrm{d}}F(X^{\mathrm{d}}) - \int_0^{X^{\mathrm{d}}} yf(y)\mathrm{d}y\right] \\
&\quad + p\int_0^{\infty} kf(X^{\mathrm{d}}-k)\left[1-\hat{F}^{n-1}\left(X^{\mathrm{d}}+\frac{k}{n-1}\right)\right]\mathrm{d}k \\
&\quad + p\int_0^{\infty} kf(X^{\mathrm{d}}+k)\hat{F}^{n-1}\left(X^{\mathrm{d}}-\frac{k}{n-1}\right)\mathrm{d}k \\
&= (r-c)X^{\mathrm{d}} - (r-v)\left[X^{\mathrm{d}}F(X^{\mathrm{d}}) - \int_0^{X^{\mathrm{d}}} yf(y)\mathrm{d}y\right] + p\int_0^{\infty} kf(X^{\mathrm{d}}-k)\mathrm{d}k \\
&= t\int_0^{X^{\mathrm{d}}} yf(y)\mathrm{d}y \\
&= t\xi(X^{\mathrm{d}}).
\end{aligned} \tag{4.14}
$$

By substituting (4.12) and (4.14) into (4.11), we obtain

$$\delta_\infty^* = \frac{\rho(X^{\mathrm{d}})}{\rho(X^{\mathrm{d}}) + t\xi(X^{\mathrm{d}}) - (r-v)\xi(X^1)}.$$

Finally, if $X^{\mathrm{d}} = \inf\{x : F(x) > \frac{r-c-p}{t}\} > \mu$, we have $\lim_{n\to\infty} 1 - \hat{F}^{n-1}(X^{\mathrm{d}} + \frac{k}{n-1}) = 0$ and $\lim_{n\to\infty} \hat{F}^{n-1}(X^{\mathrm{d}} - \frac{k}{n-1}) = 1$, hence

$$\begin{aligned}
J_i^{\mathrm{d}}(\mathbf{X}^{\mathrm{d}}, n) &= (r-c)X^{\mathrm{d}} - (r-v)\left[X^{\mathrm{d}}F(X^{\mathrm{d}}) - \int_0^{X^{\mathrm{d}}} yf(y)\mathrm{d}y\right] \\
&\quad + p\int_0^\infty kf(X^{\mathrm{d}} - k)\left[1 - \hat{F}^{n-1}\left(X^{\mathrm{d}} + \frac{k}{N-1}\right)\right]\mathrm{d}k \\
&\quad + p\int_0^\infty kf(X^{\mathrm{d}} + k)\hat{F}^{n-1}\left(X^{\mathrm{d}} - \frac{k}{n-1}\right)\mathrm{d}k \\
&= (r-c)X^{\mathrm{d}} - (r-v)\left[X^{\mathrm{d}}F(X^{\mathrm{d}}) - \int_0^{X^{\mathrm{d}}} yf(y)dy\right] + p\int_0^\infty kf(X^{\mathrm{d}} + k)\mathrm{d}k \\
&= p\mu + t\int_0^{X^{\mathrm{d}}} yf(y)\mathrm{d}y \\
&= p\mu + t\xi(X^{\mathrm{d}}). \qquad (4.15)
\end{aligned}$$

By substituting (4.12) and (4.15) into (4.11), we obtain

$$\delta_\infty^* = \frac{\rho(X^{\mathrm{d}})}{\rho(X^{\mathrm{d}}) + p\mu + t\xi(X^{\mathrm{d}}) - (r-v)\xi(X^1)}.$$

□

Proof of Proposition 5. Retailers have the same demand distribution $F(\cdot)$, price, r, salvage value, v, transshipping cost, t, and unit profit from transshipment, $p = r - v - t$. Denote $X = \sum_j X_j$, $X_{-i} = \sum_{j\neq i} X_j$ and let f^m the *p.d.f* of mD_i. It can be verified that

$$\begin{aligned}
\frac{\partial J_i^{\mathrm{d}}}{\partial X_i} - \frac{\partial J^n}{\partial X_i} &= p\int_0^\infty kf(X_i - k)f^{n-1}(X_{-i} + k)\mathrm{d}k - p\int_0^\infty kf(X_i + k)f^{n-1}(X_{-i} - k)\mathrm{d}k \\
&= p\,\mathsf{E}[X_i - D_i \,|\, X = D]\,f^n(X).
\end{aligned}$$

Denote $O_i = \left(\frac{\partial J_i^{\mathrm{d}}}{\partial X_i} - \frac{\partial J^n}{\partial X_i}\right)|_{\mathbf{X}^n}$. Achieving first best requires $O_i = 0$ for all i. However, for any $i \neq j$,

$$\begin{aligned}
O_i - O_j &= pf^n(X)\,\mathsf{E}[X_i^n - X_j^n + D_j - D_i \,|\, D = X] \\
&= pf^n(X)\left[X_i^n - X_j^n + \mathsf{E}[D_j - D_i \,|\, D = X]\right] \\
&= pf^n(X)\,(X_i^n - X_j^n).
\end{aligned}$$

It therefore requires $X_i^n = X_j^n$, $\forall i, j$. This is obviously not true given that each X_i^n has to satisfy its FOC with a different c_i:

$$\frac{\partial J^n}{\partial X_i^n} = r - c_i + (r-v)F(X_i^n) + pPr\{D_i \leq X_i^n, D > X^n\} - pPr\{D_i \geq X_i^n, D < X^n\}$$
$$= 0. \qquad \square$$

Proof of Theorem 5. The eviction contract described in Theorem 5 will be an optimal contract if it satisfies the following constraints:

1. *Participation constraint*—each retailer is better off if she adopts the contract.
2. *Early adoption constraint*—each retailer prefers to adopt the contract in the current period than in the later period.
3. *Continuation constraints*—each retailer is better off if she does not deviate in any period.

We now show that the eviction contract satisfies all three constraints.

PARTICIPATION CONSTRAINT: If retailer i adopts the contract in period 1, her infinite horizon discounted payoff is given by

$$B_i + \sum_{t=1}^{\infty} \delta_i^{t-1} J_i^n(\mathbf{X}^n) = B_i + \frac{1}{1-\delta_i} J_i^n(\mathbf{X}^n).$$

If the contract is not adopted and each retailer orders the individually optimal quantity (under the dual allocation rule), her payoff is

$$\sum_{t=1}^{\infty} \delta_i^{t-1} J_i^n(\mathbf{X}^{\mathrm{d}}) = \frac{1}{1-\delta_i} J_i^n(\mathbf{X}^{\mathrm{d}}).$$

The participation constraint is satisfied if

$$B_i + \frac{1}{1-\delta_i} J_i^n(\mathbf{X}^n) \geq \frac{1}{1-\delta_i} J_i^n(\mathbf{X}^{\mathrm{d}}).$$

First, suppose that $\Lambda_i > 0$, which implies $B_i = \frac{1}{1-\delta_i}[J_i^n(\mathbf{X}^{\mathrm{d}}) - J_i^n(\mathbf{X}^n)]$. In other words, retailer i's profit is larger if the retailers order $\mathbf{X}^{\mathrm{d}}$, and she receives a positive bonus to compensate for ordering $\mathbf{X}^n$. Then,

$$B_i + \frac{1}{1-\delta_i} J_i^n(\mathbf{X}^n) = \frac{1}{1-\delta_i}[J_i^n(\mathbf{X}^{\mathrm{d}}) - J_i^n(\mathbf{X}^n)] + \frac{1}{1-\delta_i} J_i^n(\mathbf{X}^n) = \frac{1}{1-\delta_i} J_i^n(\mathbf{X}^{\mathrm{d}}),$$

and hence i is not better off if she does not adopt the contract.

Now, suppose that $\Lambda_i \leq 0$—that is, retailer i's profit is larger if the retailers order $\mathbf{X}^n$ and she gives a side payment to other retailers to induce their acceptance of the contract. Observe that $J^n(\mathbf{X}^n) \geq J^n(\mathbf{X}^{\mathrm{d}})$, which implies $\sum_i \Lambda_i \leq 0$. This further means that $0 \leq \sum_{K^+} \Lambda_j \leq \sum_{K^-} (-\Lambda_j)$ and

$$0 \leq \frac{\sum_{K^+} (-\Lambda_j)}{\sum_{K^-} \Lambda_j} \leq 1. \tag{4.16}$$

Now,

$$\begin{aligned} B_i + \frac{1}{1-\delta_i} J_i^n(\mathbf{X}^n) &= \frac{1}{1-\delta_i}[J_i^n(\mathbf{X}^{\mathrm{d}}) - J_i^n(\mathbf{X}^n)] \times \frac{\sum_{K^+} (-\Lambda_j)}{\sum_{K^-} \Lambda_j} + \frac{1}{1-\delta_i} J_i^n(\mathbf{X}^n) \\ &\geq \frac{1}{1-\delta_i}[J_i^n(\mathbf{X}^{\mathrm{d}}) - J_i^n(\mathbf{X}^n)] + \frac{1}{1-\delta_i} J_i^n(\mathbf{X}^n) = \frac{1}{1-\delta_i} J_i^n(\mathbf{X}^{\mathrm{d}}), \end{aligned}$$

where the inequality follows from (4.16). Thus, the participation constraint is satisfied for all i.

Early adoption constraint: If the contract is adopted in period $t = 2$ instead of in period $t = 1$, the retailers order $\mathbf{X}^{\mathrm{d}}$ in period 1, and retailer i realizes the payoff

$$J_i^n(\mathbf{X}^{\mathrm{d}}) + \delta_i B_i + \sum_{t=2}^{\infty} \delta_i^{t-1} J_i^n(\mathbf{X}^n) = J_i^n(\mathbf{X}) + \delta_i B_i + \frac{\delta_i}{1-\delta_i} J_i^n(\mathbf{X}^n).$$

The early adoption constraint holds if

$$B_i + \frac{1}{1-\delta_i} J_i^n(\mathbf{X}^n) \geq J_i^n(\mathbf{X}) + \delta_i B_i + \frac{\delta_i}{1-\delta_i} J_i^n(\mathbf{X}^n).$$

First, suppose that $\Lambda_i > 0$, which implies $B_i = \frac{1}{1-\delta_i}[J_i^n(\mathbf{X}^{\mathrm{d}}) - J_i^n(\mathbf{X}^n)]$. Then,

$$\begin{aligned} J_i^n(\mathbf{X}^{\mathrm{d}}) + \delta_i B_i + \frac{\delta_i}{1-\delta_i} J_i^n(\mathbf{X}^n) &= J_i^n(\mathbf{X}^{\mathrm{d}}) + \frac{\delta_i}{1-\delta_i}[J_i^n(\mathbf{X}^{\mathrm{d}}) - J_i^n(\mathbf{X}^n)] + \frac{\delta_i}{1-\delta_i} J_i^n(\mathbf{X}^n) \\ &= \frac{1}{1-\delta_i} J_i^n(\mathbf{X}^{\mathrm{d}}), \end{aligned}$$

and

$$B_i + \frac{1}{1-\delta_i} J_i^n(\mathbf{X}^n) = \frac{1}{1-\delta_i}[J_i^n(\mathbf{X}^{\mathrm{d}}) - J_i^n(\mathbf{X}^n)] + \frac{1}{1-\delta_i} J_i^n(\mathbf{X}^n) = \frac{1}{1-\delta_i} J_i^n(\mathbf{X}^{\mathrm{d}}).$$

Hence, retailer i does not benefit from late adoption of the contract.

Next, when $\Lambda_i \leq 0$, then $J_i^n(\mathbf{X}^{\mathrm{d}}) - J_i^n(\mathbf{X}^n) \leq 0$, and (4.16) implies

$$B_i + \frac{1}{1-\delta_i} J_i^n(\mathbf{X}^n) - \left(J_i^n(\mathbf{X}^{\mathrm{d}}) + \delta_i B_i + \frac{\delta_i}{1-\delta_i} J_i^n(\mathbf{X}^n) \right)$$
$$= [J_i^n(\mathbf{X}^{\mathrm{d}}) - J_i^n(\mathbf{X}^n)] \times \frac{\sum_{K^+}(-\Lambda_j)}{\sum_{K^-}\Lambda_j} + J_i^n(\mathbf{X}^n) - J_i^n(\mathbf{X}^{\mathrm{d}})$$
$$\geq J_i^n(\mathbf{X}^{\mathrm{d}}) - J_i^n(\mathbf{X}^n) + J_i^n(\mathbf{X}^n) - J_i^n(\mathbf{X}^{\mathrm{d}}) = 0.$$

Thus, retailer i prefers to adopt the contract in the first period.

CONTINUATION CONSTRAINT: We now want to show that a retailer never benefits from defecting. Recall that Z_t denotes the coalition structure in period t, and suppose that retailer i orders a quantity different from $X_{it}^{Z_t}$ and/or withholds some of her residuals. As a result, she pays a penalty, d_{it}, in period t, and is excluded from inventory sharing in all subsequent periods. We denote, with slight abuse of notation, $\mathbf{X}_t(\hat{\mathbf{h}}_{t-1}) = \mathbf{X}_t$, $\mathbf{H}_t(\hat{\mathbf{h}}_{t-1}) = \mathbf{H}_t$, $\mathbf{E}_t(\hat{\mathbf{h}}_{t-1}) = \mathbf{E}_t$, $d_{it}(\hat{\mathbf{h}}_t) = d_{it}$, and $\Delta_{it}(\hat{\mathbf{h}}_t) = \Delta_{it}$. Then, retailer i's discounted payoff starting from period t is given by

$$J_i^{Z_t}(\mathbf{X}_t, \mathbf{H}_t, \mathbf{E}_t) + d_{it}(\mathbf{X}_t, \mathbf{H}_t, \mathbf{E}_t) + \frac{\delta_i}{1-\delta_i} J_i^1(X_i^1).$$

The continuation constraint holds if

$$J_i^{Z_t}(\mathbf{X}_t, \mathbf{H}_t, \mathbf{E}_t) + d_{it}(\mathbf{X}_t, \mathbf{H}_t, \mathbf{E}_t) + \frac{\delta_i}{1-\delta_i} J_i^1(X_i^1) \leq \frac{1}{1-\delta_i} J_i^{Z_t}(\mathbf{X}^{Z_t}).$$

If $i \in I_t^-$, then $\Delta_{it} \leq 0$, and $d_{it} = \frac{1}{1-\delta_i}\left[J_i^{Z_t}(\mathbf{X}^{Z_t}) - \delta J_i^1(X_i^1)\right] - J_i^{Z_t}(\mathbf{X}_t, \mathbf{H}_t, \mathbf{E}_t)$. Thus, i receives a payoff

$$J_i^{Z_t}(\mathbf{X}_t, \mathbf{H}_t, \mathbf{E}_t) + \frac{1}{1-\delta_i}\left[J_i^{Z_t}(\mathbf{X}^{Z_t}) - \delta_i J_i^1(X_i^1)\right] - J_i^{Z_t}(\mathbf{X}_t, \mathbf{H}_t, \mathbf{E}_t) + \frac{\delta_i}{1-\delta_i} J_i^1(X_i^1)$$
$$= \frac{1}{1-\delta_i} J_i^{Z_t}(\mathbf{X}^{Z_t}),$$

and i does not benefit from defection.

Now, suppose $i \in I_t^+$, and consequently $\Delta_{it} > 0$. This implies

$$\frac{1}{1-\delta_i}\left[J_i^{Z_t}(\mathbf{X}^{Z_t}) - \delta_i J_i^1(X_i^1)\right] - J_i^{Z_t}(\mathbf{X}_t, \mathbf{H}_t, \mathbf{E}_t) \geq 0. \tag{4.17}$$

Notice that

$$\sum_i \Delta_{it} = \sum_i \left\{ \frac{1}{1-\delta_i}\left[J_i^{Z_t}(\mathbf{X}^{Z_t}) - \delta_i J_i^1(X_i^1)\right] - J_i^{Z_t}(\mathbf{X}_t, \mathbf{H}_t, \mathbf{E}_t) \right\}$$
$$= \frac{\delta_i}{1-\delta_i}\left\{J^{Z_t}(\mathbf{X}^{Z_t}) - J^1(\mathbf{X}_1)\right\} + J^{Z_t}(\mathbf{X}^{Z_t}) - J^{Z_t}(\mathbf{X}_t, \mathbf{H}_t, \mathbf{E}_t) \geq 0,$$

where the inequality holds because $\mathbf{X}^{Z_t}$ with complete residual sharing maximizes the system profit when the state is Z_t and systems with inventory-sharing retailers generate higher profit than systems without inventory sharing. As a result, $\sum_{I_t^+} \Delta_{jt} \geq \sum_{I_t^-} (-\Delta_{jt})$, and

$$0 \leq \frac{\sum_{I_t^-} (-\Delta_{jt})}{\sum_{I_t^+} \Delta_{jt}} \leq 1. \tag{4.18}$$

Thus, retailer i receives a payoff

$$J_i^{Z_t}(\mathbf{X}_t, \mathbf{H}_t, \mathbf{E}_t) + \left\{ \frac{1}{1-\delta_i} \left[J_i^{Z_t}(\mathbf{X}^{Z_t}) - \delta_i J_i^1(X_i^1) \right] - J_i^{Z_t}(\mathbf{X}_t, \mathbf{H}_t, \mathbf{E}_t) \right\} \times \frac{\sum_{I_t^-} (-\Delta_{jt})}{\sum_{I_t^+} \Delta_{jt}}$$

$$+ \frac{\delta_i}{1-\delta_i} J_i^1(X_i^1) \leq J_i^{Z_t}(\mathbf{X}_t, \mathbf{H}_t, \mathbf{E}_t) + \frac{1}{1-\delta_i} \left[J_i^{Z_t}(\mathbf{X}^{Z_t}) - \delta_i J_i^1(X_i^1) \right] - J_i^{Z_t}(\mathbf{X}_t, \mathbf{H}_t, \mathbf{E}_t)$$

$$+ \frac{\delta_i}{1-\delta} J_i^1(X_i^1) = \frac{1}{1-\delta_i} J_i^{Z_t}(\mathbf{X}^{Z_t}),$$

where the inequality follows from (4.17) and (4.18). As a result, i prefers not to defect in any period. □

References

Anupindi, R., Bassok, Y., & Zemel, E. (2001). A general framework for the study of decentralized distribution system. *M&SOM*, *3*(4) 349–368.

Bagwell, K., & Staiger, R. W. (1997). Collusion over the business cycle. *The RAND Journal of Economics*, *28*(1) 82–106.

Dong, L., & Rudi, N. (2004). Who benefits from transshipment? Exogenous vs. endogenous wholesale prices. *Management Science*, *50*(5) 645–657.

Granot, D., & Sošić, G. (2003). A three stage model for a decentralized distribution system of retailers. *Operations Research*, *51*(5) 771–784.

Haltiwanger, J., & Harrington, J. E., Jr (1991). The impact of cyclical demand movements on collusive behavior. *The RAND Journal of Economics*, *22*(1) 89–106.

Hu, X., Duenyas, I., & Kapuscinski, R. (2007). Existence of coordinating transshipment prices in a two-location inventory problem. *Management Science*, *53*(8) 1289–1302.

Huang, X., & Sošić, G. (2010a). Transshipment of inventories: dual allocation vs. transshipment prices. *M&SOM*, *12*(2) 299–312.

Lippman, S. A., & McCardle, K. F. (1997). The competitive newsboy. *Operations Research*, *45*(1) 54–65.

Parlar, M. (1988). Game theoretic analysis of the substitutable product inventory problem with random demands. *Naval Research Logistics*, *35* 397–409.

Rotemberg, J. J., & Saloner, G. (1986). A supergame-theoretic model of price wars during booms. *The American Economic Review*, *76*(3) 390–407.

Rudi, N., Kapur, S., & Pyke, D. F. (2001). A two-location inventory model with transshipment and local decision making. *Management Science*, *47*(12) 1668–1680.

Shao, J., Krishnan, H., & McCormick, S. T. (2011). Incentives for transshipment in a supply chain with decentralized retailers. *MSOM*, *13*(3):361–372.

Sošić, G. (2004). Transshipment of inventories among retailers: myopic vs. farsighted stability. *Management Science*, *52*(10) 1493–1508.

Wang, Q., & Parlar, M. (1994). A three-person game theory model of the substitutable product inventory problem with random demands. *European Journal of Operational Research*, *76*(1) 83–97.

Wee, K. E., & Dada, M. (2005). Optimal policies for transshipping inventory in a retail network. *Management Science*, *51*(10) 1519–1533.

Yang, H., & Schrage, L. (2009). Conditions that cause risk pooling to increase inventory. *European Journal of Operational Research*, *192*(3) 837–851.

Zhang, J. (2005). Transshipment and its impact on supply chain members' performance. *Management Science*, *51*(10) 1534–1539.

Zhao, H., Deshpande, V., & Ryan, J. K. (2005). Inventory sharing and rationing in decentralized dealer networks. *Management Science*, *51*(4) 531–547.

Chapter 5
Cooperative Newsvendor Games: A Review

Luigi Montrucchio, Henk Norde, Ulaş Özen, Marco Scarsini, and Marco Slikker

Abstract In this survey, we review some of the main contributions to the cooperative approach of newsvendor situations. We show how newsvendor situations with several retailers can be modeled as a transferable-utility cooperative game and we concentrate on one solution concept: the core. First, we examine the basic model and then we consider several variations that are of interest from a theoretical and an applied viewpoint.

Keywords Newsvendor model • Core • Shapley value • Large games • Multiple warehouses • Transshipment costs • Stochastic programming • Duality

---

L. Montrucchio
Dipartimento di Statistica e Matematica Applicata and Collegio Carlo Alberto, Università di Torino, Corso Unione Sovietica, 218/bis, I–10134 Torino, Italy
e-mail: luigi.montrucchio@unito.it

H. Norde
Center and Department of Econometrics and OR, Tilburg University, 5000 LE Tilburg, Netherlands
e-mail: h.norde@uvt.nl

U. Özen
Alcatel-Lucent, Bell Labs Ireland, Blanchardstown Industrial Park, Dublin 15, Ireland
e-mail: ulas.ozen@alcatel-lucent.com

M. Scarsini (✉)
Dipartimento di Economia e Finanza, LUISS, Viale Romania 12, I–00197 Roma, Italy
e-mail: marco.scarsini@luiss.it

M. Slikker
School of Industrial Engineering, Technische Universiteit Eindhoven, Paviljoen E13, 5600 MB Eindhoven, Netherlands
e-mail: M.Slikker@tue.nl

T.-M. Choi (ed.), *Handbook of Newsvendor Problems*, International Series in Operations Research & Management Science 176, DOI 10.1007/978-1-4614-3600-3_5,

5.1 Introduction

In recent years, many companies have started to employ cooperation strategies to improve their service offerings, reduce costs, and increase profit margins. Several collaborative actions have been reported in various industries including pharmaceutical, fashion, automobile, and post-sales support (see Anupindi and Bassok 1999; Chen and Zhang 2009; Chen 2009, for their motivating examples). The main driver of these collaborative activities is the cost benefit due to economies of scale and risk pooling, which can be achieved in several operations including logistics, purchasing, and inventories. Among them, inventories have a special importance because in many supply chains they constitute a big portion of the overall investment and resources. Benefits of physical pooling of inventories have been long studied mainly for intercompany operations. Recent developments in information technologies and logistic networks allowed firms to further exploit the benefit of physical pooling of inventories by virtual pooling in which the stocks are kept locally but can be transferred to satisfy demand (see Anupindi et al. 2001). This opened new opportunities for independent firms to reduce inventory-related costs by cooperation and (physical or virtual) inventory centralization. One typical example is seen in the automobile industry. Car dealers often cooperate by transferring stocks if a fellow dealer cannot satisfy a customer's request from his/her own stock. This way, they can increase revenues as well as improve customer satisfaction.

Before any collaboration can be established, probably the first decision that firms need to make is the allocation of anticipated costs and benefits among themselves. This allocation should be considered advantageous by all the firms to motivate them to cooperate. Finding such an allocation might be a nontrivial task even though collaboration usually improves total costs and revenues. In a two-party situation, the answer would be easy. Any allocation improving the firms' stand-alone profit will be considered as a win–win case. However, if more firms are involved, they want to know that it is not more advantageous for them to cooperate in a smaller group rather than joining the big group. Existence of such an allocation is crucial for the stability of the cooperation. This problem can be studied using cooperative game theory (also called coalitional game theory), which provides, among others, the core as a popular solution concept. The core is the set of allocations upon which no coalition can improve. Depending on the characteristics of the game, core allocations may exist or may fail to exist, i.e., the core can be empty. In this survey, we review the papers studying inventory cooperation using newsvendor models in combination with the core as solution concept for the allocation problem of anticipated benefits.

The newsvendor model is probably the most celebrated model in inventory literature and it is used extensively to study inventory centralization and cooperation. Newsvendor models are used especially for products with high perishability or short life cycles. Initial applications appeared in the fashion industry and were then extended to other industries with decreasing life cycles, such as high tech. Moreover, because of their ability to capture the basic tradeoffs regarding inventory-related decisions, newsvendor models and their extensions became backbone models to study effects of inventory centralization—namely cost reduction and profit increase—in many decentralized supply chain systems. Inventory cooperation and allocation of benefits has been a subject of inventory literature for a long time. Early papers studied slightly different models and approaches to the problem. Eppen (1979) was one of the first to consider the effect of centralization for a multilocation newsvendor problem. His approach is not game theoretic. Parlar (1988) used noncooperative game theory to analyze a simple inventory model with two retailers and studied what happens when they decide to cooperate. Wang and Parlar (1994) considered a three-player noncooperative model and provided conditions for cooperation. Gerchak and Gupta (1991) proposed different cost allocation rules in centralized inventory systems. In response to their article, Robinson (1993) proposed the use of the Shapley value to allocate costs.

In the last decade, cooperative game theory and the core concept became a mean for studying this problem. This part of the literature is the focus of our review paper. Hartman (1994) introduced the newsvendor centralization game that will be described in Sect. 5.3. Hartman and Dror (1996) proposed three desirable criteria for cost allocation rules: stability (core of a related cooperative game), justifiability (consistency of benefits with costs), and polynomial computability. They showed that common allocation procedures may fail to satisfy all three criteria, and they proposed a rule that meets all three. Hartman et al. (2000) showed that the newsvendor game is balanced (i.e., the core of such a game is nonempty) in some interesting cases, such as normally distributed demands. Müller et al. (2002) showed that any finite newsvendor game is balanced whenever the random demand vectors has finite mean. This result was independently obtained by Slikker et al. (2001) and generalized to other settings by several authors. For instance, Slikker et al. (2005) proved balancedness for games with transshipment costs. Montrucchio and Scarsini (2007) proved that the core is nonempty also for newsvendor games with infinitely many players. Özen et al. (2008) dealt with balancedness for games with several warehouses. Chen and Zhang (2009) developed a stochastic programming duality approach to find a core element of these games.

Game theoretical analysis of inventory centralization, supply chain cooperation as well as competition is extensive in the literature. The reader is referred to Cachon and Netessine (2004) and Leng and Parlar (2005) for a comprehensive survey of applications of game theory to supply chain management. For a review of inventory centralization games of deterministic and stochastic models, see Fiestras-Janeiro et al. (2011). Dror and Hartman (2011) extended this review with a specific focus on joint replenishment games and newsvendor realization games. For a comprehensive review of game theoretical models of supply chain management, see Nagarajan and Sošić (2008).

The paper is organized as follows. In Sect. 5.2, we introduce some basic concepts in cooperative and noncooperative game theory. In Sect. 5.3, the fundamental results for the basic newsvendor game are presented. Section 5.4 deals with several extensions, such as large games, games with multiple warehouses, nonlinear costs, etc.

5.2 Preliminaries on Game Theory

5.2.1 Cooperative Games

In this section we provide some basic concepts of cooperative game theory. We refer the reader to Peleg and Sudhölter (2007) for an extended treatment of this subject.

Consider a finite set $N = \{1, \dots, n\}$ of players and a function $v : 2^N \to \mathbb{R}$, called *characteristic function*, such that $v(\varnothing) = 0$. We call $\mathcal{G} = \langle N, v\rangle$ a (finite) *cooperative game with transferable utility* (or simply TU game). A subset $S \subseteq N$ is a coalition and $v(S)$ is the worth that coalition S can achieve by itself. The worth of the coalition can be transferred among its players and this justifies the name. The set N is often called the *grand coalition*. When there is no risk of confusion we will denote the game by v rather than $\langle N, v\rangle$.

Given a coalition S, the *subgame* $\langle S, v_S\rangle$ is the game with grand coalition S and characteristic function v_S such that $v_S(T) = v(T)$ for $T \subseteq S$.

An allocation is a vector $\mu \in \mathbb{R}^N$ such that $\sum_{i\in N} \mu_i = v(N)$. Such a vector is a possible way to split the worth of the grand coalition among all players efficiently, that is, without any leftover.

The *core* of the game $\langle N, v\rangle$ is defined as

$$\mathrm{core}(\langle N, v\rangle) := \left\{ \mu \in \mathbb{R}^N : \sum_{i\in S} \mu_i \geq v(S) \text{ for all } S \subset N, \text{ and } \sum_{i\in N} \mu_i = v(N) \right\}.$$

If an allocation is not in the core, then it is not stable, because there exists a coalition S such that $v(S) > \sum_{i\in S} \mu_i$. This coalition will therefore have an incentive to deviate and achieve $v(S)$ by itself rather than join the grand coalition and obtain only $\sum_{i\in S} \mu_i$.

Call e^i the i-th vector of the canonical basis of $\mathbb{R}^n$ and define $e^S := \sum_{i\in S} e^i$. A map $\kappa : 2^N \to [0, 1]$ is called *balanced* if

$$\sum_{S\in 2^N\setminus\{\varnothing\}} \kappa(S) e^S = e^N.$$

A game (N, v) is called *balanced* if for every balanced map κ, we have

$$\sum_{S\in 2^N} \kappa(S) v(S) \leq v(N).$$

The above condition considers the situations where the players can form subcoalitions (i.e., every balanced map represents a situation where each player i forms coalition S with $i \in S$ $\kappa(S)$ fraction of his time), and checks whether the players, if they organized themselves via these subcoalitions with corresponding weights, can do better than the grand coalition.

Theorem 5.2.1 (Bondareva 1963; Shapley 1967). *Let* $\langle N, v\rangle$ *be a TU game. Then,* $\mathrm{core}(\langle N, v\rangle) \neq \varnothing$ *if and only if* (N, v) *is balanced.*

A game $\langle N, v\rangle$ is called *totally balanced* if it is balanced and each of its subgames is balanced as well.

Another solution concept for cooperative games was proposed by Shapley (1953). Consider a game $\langle N, v\rangle$ and a bijection $\sigma : N \to \{1,\ldots,n\}$. If players arrive in the order $\sigma(1), \sigma(2), \ldots, \sigma(n)$ (i.e., player i arrives on position $\sigma(i)$), denote by $p_\sigma(i)$ the set of players that precede i in σ, i.e.,

$$p_\sigma(i) = \{j \in N : \sigma(j) < \sigma(i)\},$$

and call $m^i_\sigma(v)$ the marginal contribution of player i, i.e.,

$$m^i_\sigma(v) = v(p_\sigma(i) \cup \{i\}) - v(p_\sigma(i)).$$

The *Shapley value* ϕ of the game $\langle N, v\rangle$ assigns to player i her average marginal contribution, where the average is taken over all possible permutations σ:

$$\phi^i(v) = \sum_\sigma \frac{1}{n!} m^i_\sigma(v).$$

It can be shown that the above expression can be written also as

$$\phi^i(v) = \sum_{S \subset N : i \notin S} \frac{|S|!(n-1-|S|)!}{n!} (v(S \cup \{i\}) - v(S)).$$

A game $\langle N, v\rangle$ is called *convex* (or *supermodular*) if

$$v(S) + v(T) \leq v(S \cap T) + v(S \cup T) \text{ for all } S, T \subseteq N. \tag{5.1}$$

It can be proved that for the case considered here of finite games (5.1) is equivalent to

$$v(T \cup \{i\}) - v(T) \geq v(S \cup \{i\}) - v(S) \text{ for all } i \in N \text{ and all } S \subset T \subseteq N \setminus \{i\}. \tag{5.2}$$

Hence, for convex games, the marginal contribution of any player to any coalition is greater than her marginal contribution to a smaller coalition. Convex games are well known for having several nice properties related to the structure of the core and solution concepts. For example, Shapley (1971) and Ichiishi (1990) showed that the marginal vectors of a game are the extreme points of the core if and only if the game is convex. Moreover, the Shapley value of a convex game is the barycenter of its core and hence it is always in the core.

In Sect. 5.4.1, we will consider games with infinitely (possibly uncountably) many players. This more general setting requires a richer structure and some technical details have to be settled precisely. A cooperative game is now defined as $\mathcal{G} = \langle I, \mathcal{C}, v \rangle$, where I is the set of players, $\mathcal{C}$ is a σ-algebra of subsets of I (the feasible coalitions), and $v : \mathcal{C} \to \mathbb{R}$ is a function such that $v(\varnothing) = 0$. To emphasize the difference between finite and infinite games, we now use the symbol I rather than N to denote the set of players. Note that with $I = N$ and $\mathcal{C} = 2^N$, this definition covers the definition of a finite cooperative game as well.

A map $\mu : \mathcal{C} \to \mathbb{R}$ is called *additive* if $\mu(A \cup B) = \mu(A) + \mu(B)$ for every $A, B \in \mathcal{C}$ with $A \cap B = \varnothing$ and *countably additive* if $\mu(\cup_{i=1}^{\infty} A_i) = \sum_{i=1}^{\infty} \mu(A_i)$ for every $A_1, A_2, \ldots \in \mathcal{C}$ with $A_i \cap A_j = \varnothing$ when $i \neq j$.

The concept of allocation in the finite setting is generalized in a straightforward way to the infinite setting by considering bounded additive maps $\mu : \mathcal{C} \to \mathbb{R}$. The collection of these maps is denoted by $\mathrm{ba}(\mathcal{C})$. The *core* of a game v is the set

$$\mathrm{core}(v) = \{\mu \in \mathrm{ba}(\mathcal{C}) : \mu(S) \geq v(S) \text{ for all } S \in \mathcal{C} \text{ and } \mu(I) = v(I)\}.$$

The game v is *exact* if its core is nonempty and

$$v(S) = \min_{\mu \in \mathrm{core}(v)} \mu(S), \qquad \text{for every } S \in \mathcal{C}.$$

Exact games, introduced by Schmeidler (1972), are a special subclass of the class of totally balanced games. They have the property that for every coalition S there is an element μ in the core such that $\mu(S) = v(S)$. This implies in turn that its restriction μ_S lies in the core of the subgame $\langle S, v_S \rangle$. Clearly this property gives rise to a tight connection between the form of the game and its core. Note further that convex games are obviously exact. Therefore, exact games can also be viewed as a intermediate class between totally balanced games and convex ones.

A version of Theorem 5.2.1 holds also for positive infinite games, as proved by Schmeidler (1967) and Kannai (1969). An extension to bounded (not necessarily positive) games can be found in Marinacci and Montrucchio (2004, Theorem 4.1).

5.2.2 *Noncooperative Games*

In this section, we provide some basic concepts of noncooperative game theory that will be used in Sect. 5.4.4.

A *game in strategic form* is a tuple $\Lambda = \langle N, Y, u \rangle$ where $N = \{1, \ldots, n\}$ is a finite set of players, $Y = \times_{i \in N} Y_i$ is the Cartesian product of strategy sets Y_i of player $i \in N$ and $u = (u_i)_{i \in N}$ is the vector of payoff functions $u_i : Y \to \mathbb{R}$ of player $i \in N$. In this game, every player i selects a strategy $y_i \in Y_i$ simultaneously. As a result of these strategic decisions, each player i receives a payoff $u_i(y)$ where $y = (y_i)_{i \in N}$ is the strategy profile played by the players. For any $S \subseteq N$ and any strategy profile $y \in Y$ we will write $y_S = (y_i)_{i \in S}$ and $y = (y_S, y_{N \setminus S})$. A strategy profile $y^* \in Y$ is a *Nash Equilibrium* if $u_i(y^*) \geq u_i(y_i, y^*_{N \setminus \{i\}})$ for all $i \in N$ and all $y_i \in Y_i$. Under a

Nash equilibrium strategy, none of the players can increase his utility by a unilateral deviation in his strategy. A strategy profile $y^* \in Y$ is called a strong Nash equilibrium (see Aumann 1959) if there is no $T \subseteq N$ and $y_T = (y_i)_{i \in T} \in \times_{i \in T} Y_i$ such that $u_i(y_T, y^*_{N \setminus T}) \geq u_i(y^*)$ for all $i \in T$ with the inequality being strict for at least one player $i \in T$. Therefore, under a strong Nash equilibrium, no group of players can increase their payoffs simultaneously by deviating collectively. The set of all strong Nash equilibria of Λ is denoted by $\mathscr{S}(\Lambda)$.

5.3 The Basic Model

5.3.1 The Newsvendor Problem

We introduce the newsvendor problem in an abstract setting that will prove suitable for the analysis of the game. Let $(\Omega, \mathscr{F}, \mathbb{P})$ be a probability space, i.e., Ω is the set of states of the world, $\mathscr{F}$ is a σ-algebra of subsets of Ω and $\mathbb{P}$ is a probability measure on $\mathscr{F}$. All random variables in this paper are assumed to be integrable, i.e., elements of $L^1(\Omega, \mathscr{F}, \mathbb{P})$. In case there is no ambiguity about the probability space $(\Omega, \mathscr{F}, \mathbb{P})$, we write L^1 instead of $L^1(\Omega, \mathscr{F}, \mathbb{P})$.

A newsvendor has to decide how many newspapers to stock in order to face an unknown demand, knowing that no replenishment is allowed. If she faces a demand x and orders a quantity y, then she obtains a profit

$$\psi(x,y) = p\min\{x,y\} - cy = \begin{cases} px - cy & \text{if} \quad x \leq y, \\ (p-c)y & \text{if} \quad x > y, \end{cases} \tag{5.3}$$

where c is the unitary cost that the newsvendor incurs for buying a newspaper from the publisher and p is the unitary gain for selling a newspaper to a customer. Obviously, $p > c$.

If the newsvendor faces random demand $X \in L^1$ she has to find an order quantity y that maximizes $\mathbb{E}[\psi(X,y)]$. The operator $\Pi : L^1 \to \mathbb{R}$, defined by

$$\Pi_\psi(X) = \max_{y \in \mathbb{R}} \mathbb{E}[\psi(X,y)], \tag{5.4}$$

represents the expected profit for a newsvendor who orders the optimal amount of newspapers.

It is not difficult to prove that the maximizer in (5.4) is a $(p-c)/p$-quantile of the distribution of X, that is $y^* \in \arg\max \mathbb{E}[\psi(X,y)]$ if

$$F_X(s) \leq \frac{p-c}{p} \leq F_X(t), \quad \text{for all } s \leq y^* \leq t, \tag{5.5}$$

where F_X is the distribution function of X. Therefore, $\Pi_\psi(X)$ can be rewritten as

$$\Pi_\psi(X) = (p-c)y^* - p\int_{u\le y^*} F_X(u)\mathrm{d}u. \tag{5.6}$$

5.3.2 The Simple Newsvendor Game

In this section, we introduce the simplest model of newsvendor games. In order to do so, we first introduce a *newsvendor situation* as $\Gamma = \langle N, (X_i)_{i\in N}, c, p\rangle$, where:

1. $N = \{1,\dots,n\}$ is the set of retailers
2. $X_i \in L^1$ is the random demand of retailer i
3. c is the unitary cost for ordering from the warehouse
4. p is the unitary selling price for the retailer

For a coalition $S \subseteq N$, let $X(S) := \sum_{i\in S} X_i$ be the aggregate random demand. The expected optimal profit of S is $\Pi_\psi(X(S))$, where Π_ψ is as defined in (5.4). So all coalitions have the same anonymous profit function and face the same type of maximization problem, the only difference being the random demand that they face. Define $v^\Gamma : 2^N \to \mathbb{R}$ as follows:

$$v^\Gamma(S) = \Pi_\psi(X(S)) \text{ for all } S \subseteq N.$$

The game $\langle N, v^\Gamma\rangle$ is called the *newsvendor game*, corresponding to newsvendor situation Γ.

Theorem 5.3.1 (Müller et al. 2002; Slikker et al. 2001). *Every newsvendor game has a nonempty core.*

This result was obtained independently in the two papers. Müller et al. (2002) actually considered costs instead of profits. In fact, instead of (5.3) they considered the function

$$\chi(x,y) = \begin{cases} c(y-x) & \text{if} \quad x \le y, \\ (p-c)(x-y) & \text{if} \quad x > y, \end{cases}$$

and referred to c as "unit holding cost" and $p-c$ as "unit penalty cost for unsatisfied demand". Subsequently, they defined the cost game $c^\Gamma : 2^N \to \mathbb{R}$ by:

$$c^\Gamma(S) = \Pi_\chi(X(S)) \text{ for all } S \subseteq N.$$

It is straightforward to check that

$$v^\Gamma(S) = \sum_{i\in S}(p-c)\mathbb{E}(X_i) - c^\Gamma(S)$$

for all $S \subseteq N$. That is, the two games differ by a factor -1 and an additive game; hence, from a game theoretical point of view, they are similar. As recognized by Slikker et al. (2001), substitution of variables combined with the addition of an appropriate additive game can prove useful in incorporating additional features while keeping the game under investigation strategically equivalent to the original game. In this way, Montrucchio and Scarsini (2007) considered a setting with unitary gains, unitary costs as well as penalty costs for not ordering enough.

As each subgame is a newsvendor game in itself, total balancedness, i.e., nonempty cores for the game and all its subgames, follows naturally. A similar remark holds for other results in this chapter as well.

The following result determines an allocation that is always contained in the core.

Theorem 5.3.2 (Chen and Zhang 2009; Montrucchio and Scarsini 2007). *If Γ is a newsvendor situation such that X_i has a continuous distribution for every $i \in N$, then $\mu \in \mathrm{core}(\langle N, \nu^\Gamma \rangle)$, where*

$$\mu_i = p \int_{X(N) \leq y^*} X_i \, \mathrm{d}\mathbb{P}$$

for every $i \in N$ and y^* is a $(p-c)/p$-quantile of the distribution of $X(N)$.

Montrucchio and Scarsini (2007) proved the result constructively for general newsvendor games (not necessarily finite). Chen and Zhang (2009) used stochastic programming duality to find an allocation in the core of a large class of inventory games (see Sect. 5.4.3). For the basic newsvendor game the two approaches give the same result.

Notice that in Theorem 5.3.1 the only assumption is that $X_i \in L^1$ for every $i \in N$, in particular no assumption is made about the dependence structure of $(X_i)_{i \in N}$. When some assumptions are made, stronger results can be obtained. This happens, for instance, when the vector $(X_i)_{i \in N}$ is comonotonic. The reader is referred, for instance, to Puccetti and Scarsini (2010) for results and references about comonotonicity.

Definition 5.3.3. Given a random vector $X = (X_i)_{i \in N}$, its *support* $\mathrm{supp}(X)$ is the smallest closed set $A \subset \mathbb{R}^n$ such that $\mathbb{P}(X \in A) = 1$. A random vector $(X_i)_{i \in N}$ is *comonotonic* if $\mathrm{supp}(X)$ is totally ordered, i.e., for all $x, y \in \mathrm{supp}(X)$, either $x \leq y$ or $y \leq x$ (in the natural component-wise order on $\mathbb{R}^n$).

Proposition 5.3.4 (Müller et al. 2002). *If Γ is a newsvendor situation such that $(X_i)_{i \in N}$ is comonotonic, then $\mathrm{core}(\langle N, \nu^\Gamma \rangle)$ is a singleton.*

When the demands of all the retailers are comonotonic, they all move in the same direction, which implies that no hedging is possible. Proposition 5.3.4 shows that in this case the core consists of only one possible allocation of profits.

Proposition 5.3.5 (Müller et al. 2002). *If Γ is a newsvendor situation such that for all $S \subseteq N$ and for all $i \in S$ the function $s \mapsto \mathbb{E}[X_i | X(S) = s]$ is increasing, then for all $\mu \in \mathrm{core}(\langle N, \nu^\Gamma \rangle)$, we have $\mu_i \leq (p-c)\mathbb{E}(X_i)$ for all $i \in N$.*

In the cost model studied by Müller et al. (2002) Proposition 5.3.5 has a more readable meaning. It just says that if $\mathbb{E}[X_i|X(S)=s]$ is increasing in s, then all cost allocations in the core are positive.

Definition 5.3.6. A random vector $(X_i)_{i\in N}$ is *exchangeable* if its distribution function is invariant with respect to permutations, i.e., for every permutation π of $\{1,\dots,n\}$ and every $(x_1,\dots,x_n)\in\mathbb{R}^n$, we have

$$\mathbb{P}(X_1 \le x_1,\dots,X_n \le x_n) = \mathbb{P}(X_1 \le x_{\pi(1)},\dots,X_n \le x_{\pi(n)}).$$

Proposition 5.3.7 (Müller et al. 2002). *If Γ is a newsvendor situation such that for some vector $a\in\mathbb{R}^N$ the random vector $(X_i - a_i)_{i\in N}$ is exchangeable, then there exists a $\mu \in \mathrm{core}(\langle N, v^\Gamma\rangle)$ such that $\mu_i \le (p-c)\mathbb{E}(X_i)$ for all $i\in N$.*

Note that in Propositions 5.3.5 and 5.3.7 the statement that "$\mu \in \mathrm{core}(\langle N, v^\Gamma\rangle)$ is such that $\mu_i \le (p-c)\mathbb{E}(X_i)$ for all $i\in N$" is equivalent to the statement that the corresponding element in the (anti-)core of $\langle N, c^\Gamma\rangle$ is nonnegative.

As we mentioned in Sect. 5.2, convexity is an important property of cooperative games. Regarding newsvendor games, convexity suggests that the marginal benefit of inventory pooling is increasing with the size of the coalition. Though newsvendor games have nonempty cores, in general they are not convex. Counterexamples can be found even in very simple settings (see, for instance, Özen et al. 2011, Examples 3.1 and 3.2). The demand distribution of the retailers and the optimal fractile are two important factors that affect the convexity of newsvendor games. Özen et al. (2011, Examples 6.1) showed that even newsvendor games with independent, symmetric, and unimodal demand distributions need not be convex.

The following proposition provides sufficient conditions for the convexity of a newsvendor game.

Proposition 5.3.8 (Montrucchio and Scarsini 2007; Özen et al. 2011; Slikker et al. 2001). *The newsvendor game $\langle N, v^\Gamma\rangle$ is convex if one of the following conditions holds:*

1. *$(X_i)_{i\in N}$ are independent, symmetric, and unimodal distributed random variables and $c = p/2$.*
2. *$(X_i)_{i\in N}$ are independent normal distributed random variables.*
3. *$(X_i)_{i\in N}$ are exchangeable multi-normal distributed random variables.*

Proposition 5.3.8 states that under some realistic assumptions the newsvendor game turns out to be convex, a very nice property of TU games as we discuss in Sect. 5.2.1. First of all, these games are convex if demand is independent, symmetric, and unimodal distributed and $c = p/2$, which means that the retailers profit margin is 100%. Secondly, they satisfy convexity under independent normal demand distributions regardless of the profit margin that the retailers work with. Finally, convexity extends to situations with identical and positively correlated normal demand distributions.

5.4 Extensions

5.4.1 Large Games

In many situations, it is useful to consider newsvendor games with a large number of retailers. These games are often computationally complex, whereas an infinite approximation of them can be easier to handle. This is why Montrucchio and Scarsini (2007) considered newsvendor games where the set of retailers is large, possibly uncountable. In order to define these games properly we first adjust the concept of newsvendor situations to this setting. A *large newsvendor situation* is a tuple $\Gamma^l = \langle (I,\mathscr{C}),(X_S)_{S\in\mathscr{C}},c,p\rangle$, where:

1. I is a (possibly uncountable) set of retailers and $\mathscr{C}$ is a σ-algebra of subsets of I.
2. $X_S \in L^1$ is the joint random demand of coalition $S \in \mathscr{C}$; for all $\omega \in \Omega$, the map $S \mapsto X_S(\omega)$ is additive on $\mathscr{C}$, that is, $X_{S_1\cup S_2}(\cdot) = X_{S_1}(\cdot) + X_{S_2}(\cdot)$ for two disjoint coalitions $S_1, S_2 \in \mathscr{C}$.
3. c is the unitary cost for ordering from the warehouse.
4. p is the unitary selling price for the retailer.

Note that with $I = N$, $\mathscr{C} = 2^N$, and $X_S = \sum_{i\in S} X_i$ for all $S \subseteq N$, this definition covers the newsvendor situations with a finite retailer set as well.

For $S \in \mathscr{C}$, consider the joint random demand X_S for coalition S. The expected optimal profit of S is $\Pi_\psi(X_S)$, where Π_ψ is again as defined in (5.4). Define v^{Γ^l} : $\mathscr{C} \to \mathbb{R}$ as follows:

$$v^{\Gamma^l}(S) = \Pi_\psi(X_S) \text{ for all } S \in \mathscr{C}.$$

The game $\langle N, v^{\Gamma^l}\rangle$ is called the *large newsvendor game*, corresponding to newsvendor situation Γ^l.

Theorem 5.4.1 (Montrucchio and Scarsini 2007). *If Γ^l is a large newsvendor situation such that*

$$\int |X_S| \, \mathrm{d}\mathbb{P} \le K \quad \textit{for all } S \in \mathscr{C} \textit{ and some } K > 0, \tag{5.7}$$

then the newsvendor game $\langle I, v^{\Gamma^l}\rangle$ is totally balanced.

Moreover, if the aggregate demand X_I has a continuous distribution, then $\mu \in$ core$(\langle I, v^{\Gamma^l}\rangle)$, where $\mu : \mathscr{C} \to \mathbb{R}$ is defined as

$$\mu(S) = p\int_{X_I \le y^*} X_S \, \mathrm{d}\mathbb{P} \tag{5.8}$$

for all $S \in \mathscr{C}$, and where y^* is a $(p-c)/p$-quantile of the distribution of X_I.

Notice that condition (5.7) is automatically satisfied when the map $S \mapsto X_S$ is countably additive.

As in the finite case, stronger results can be obtained by assuming some particular dependence structure of the random demand.

Proposition 5.4.2 (Montrucchio and Scarsini 2007). *If Γ^l is a large newsvendor situation such that X_I and X_S are comonotone for all coalitions $S \in \mathscr{C}$, then the newsvendor game $\langle I, \nu^{\Gamma^l} \rangle$ is exact.*

In many situations with many retailers, the influence of each one of them is negligible. This can be expressed by assuming that the random demand vector is nonatomic. The next theorem establishes that in such a nonatomic setting, when the aggregate demand has a continuous distribution, the core of newsvendor games is a singleton.

Theorem 5.4.3 (Montrucchio and Scarsini 2007). *If Γ^l is a large newsvendor situation such that random demand vector $(X_S)_{S \in \mathscr{C}}$ has a Radom-Nikodym derivative and such that X_I has a continuous distribution, then* $\mathrm{core}(\langle I, \nu^{\Gamma^l} \rangle)$ *is a singleton, given by (5.8).*

5.4.2 *Multiple Warehouses*

In the newsvendor games covered so far, a complete consolidation of stocks is assumed and all retailers face the same purchasing cost c and selling price p. Several extensions of this simple cost structure are considered in the literature. The model of Özen et al. (2008) presents such an extension with multiple warehouses.

Consider a distribution system with n retailers and m warehouses. Every retailer i faces a stochastic demand and sells a single product. As in the standard newsvendor problem, retailers have to decide on order quantities for one or more warehouses, before the stochastic demand is realized. After the stochastic demand is realized, stocks that are available at the warehouses are allocated to the retailers to satisfy the demand. Formally, a *newsvendor situation with warehouses* is a tuple $\Gamma^w = \langle N, W, (Z_i)_{i \in N}, (X_i)_{i \in N}, (c_w)_{w \in W}, (f_{wi})_{w \in W, i \in N}, (p_i)_{i \in N} \rangle$, where[1]

1. $N = \{1, \ldots, n\}$ is the set of retailers.
2. $W = \{1, \ldots, m\}$ is the set of warehouses.
3. $Z_i \subseteq W$ is the nonempty set of warehouses related to retailer i.
4. $X_i \in L^1$ is the random demand of retailer i.
5. $c_w > 0$ is the unitary cost for ordering via warehouse w.

[1] We remark that in Özen et al. (2008) there is no explicit unitary cost for ordering via a warehouse, as it is assumed to be incorporated in the transportation costs. With all orders being sent to the retailers, as prescribed by the admissible allocations in M^S below, the two models are equivalent. Formally, the models coincide by setting unitary transportation costs from warehouse w to retailer i in Özen et al. (2008) equal to $\overline{f}_{wi} = f_{wi} + c_w$.

6. $f_{wi} \geq 0$ is the unitary transportation cost from warehouse w to retailer i.
7. $p_i > 0$ is the unitary selling price for the retailer i.

Each retailer i can order goods from the set of warehouses Z_i. Since $Z_i \subseteq W$, this model covers two extreme cases:

(a) Every retailer orders only from one warehouse, i.e., $|Z_i| = 1$ for all $i \in N$.
(b) Every retailer can use all warehouses, i.e., $Z_i = W$ for all $i \in N$.

Forming coalition $S \subseteq N$, the retailers are allowed to use any warehouse in $Z_S := \bigcup_{i \in S} Z_i$. This generalizes several models studied earlier in the literature and covers a broad range of situations in which the reallocation of the orders can take place in different locations between the supplier and the retailers. For example, the situation with $W = \{1\}$, $Z_i = W$, and $f_{1i} = 0$ for all $i \in N$, and $p_i = p$ represents the simple newsvendor model as described in Sect. 5.3.2. Alternatively, a situation with $W = N$ and $Z_i = \{i\}$ for all $i \in N$ represents a system in which the retailers keep local stock and cooperate through lateral transshipment as studied in Slikker et al. (2005).

Consider a coalition $S \subseteq N$. Let $x^S = (x_i)_{i \in S}$ be a realization of the random vector $X^S = (X_i)_{i \in S}$. Before the realization of the random vector X^S, the retailers in the coalition jointly choose an order vector q^S from the set

$$Q^S := \{q \in \mathbb{R}^m | q_w = 0 \text{ for all } w \in W \setminus Z_S \text{ and } q_w \geq 0 \text{ for all } w \in Z_S\}.$$

The cost of placing this order is

$$C(q^S) = \sum_{w \in W} c_w q_w^S. \tag{5.9}$$

After observing the realization x^S of X^S, the players in S decide on an allocation A^S of joint orders from the set

$$M^S(q^S) := \left\{ A^S \in \mathbb{R}_+^{m \times n} | \sum_{i \in S} A_{wi}^S = q_w^S \text{ for all } w \in Z_S \right.$$
$$\left. \text{and } A_{wi}^S = 0 \text{ if } i \in N \setminus S \text{ or } w \in W \setminus Z_S \right\}.$$

The revenue created at each retailer is

$$H^i(A_i^S, x_i^S) = - \sum_{w \in Z_S} A_{wi}^S f_{wi} + p_i \min \left\{ \sum_{w \in Z_S} A_{wi}^S, x_i^S \right\}. \tag{5.10}$$

Hence, the coalition's total revenue is given by

$$R^S(A^S, x^S) = \sum_{i \in S} H^i(A_i^S, x_i).$$

Lemma 5.4.4 (Özen et al. 2008). Let Γ^w be a newsvendor situation with warehouses and let $S \subseteq N$ be a coalition. For any $q^S \in Q^S$ and demand realization vector x^S, there exists an allocation $A^{S,*} \in M^S(q^S)$ that maximizes total revenue $R^S(\cdot, x^S)$.

We refer to $R^S(A^{S,*}, x^S)$ as $r^S(q^S, x^S)$, which is the maximum total revenue that coalition S can achieve by allocating the available stocks optimally. Then, the expected profit function of coalition S is defined by

$$\pi^S(q^S) = \mathbb{E}[r^S(q^S, X^S)] - C(q^S).$$

Theorem 5.4.5 (Özen et al. 2008). *Let Γ^w be a newsvendor situation with warehouses and let $S \subseteq N$ be a coalition. There exists an order vector $q^{S,*}$ that maximizes the expected profit function $\pi^S(\cdot)$.*

Let Γ^w be a newsvendor situation with warehouses. According to Theorem 5.4.5, we can define $v^{\Gamma^w} : 2^N \to \mathbb{R}$ by

$$v^{\Gamma^w}(S) = \max_{q^S \in Q^S} \pi^S(q^S) \text{ for all } S \subseteq N.$$

The game $\langle N, v^{\Gamma^w} \rangle$ is called the *newsvendor game with warehouses*, corresponding to Γ^w.

Introducing multiple warehouses into the model, can create some externalities. Although the retailers may prefer not to use a specific warehouse if they act alone, this warehouse may be used if they cooperate, for example, because of a strategic position of this warehouse somewhere between the retailers. Moreover, the core of general newsvendor games can shrink when more warehouses are introduced into the system as shown by Özen et al. (2008, Example 3). However, the following theorem shows that the core of a newsvendor game with warehouses is never empty.

Theorem 5.4.6 (Özen et al. 2008). *Every newsvendor game with warehouses has a nonempty core.*

5.4.3 A Stochastic Programming Duality Approach

Chen and Zhang (2009) studied a cost game related to a newsvendor network with warehouses Γ^w and developed an alternative approach by formulating their game via a two-stage stochastic program. Denote $X^S(\omega) = (X_i(\omega))_{i \in S}$. Then

$$v^{\Gamma^w}(S) = \max \left(- \sum_{w \in Z^S} c_w q_w + \mathbb{E}\left[r^S(q, X^S(\cdot)) \right] \right), \tag{5.11}$$

$$\text{s.t.} \quad q_w \geq 0, \quad w \in Z^S. \tag{5.12}$$

Here, $q = (q_w)_{w \in Z^S}$. For every q and ω, we have

$$r^S(q, X^S(\omega)) = \max \left(\sum_{i \in S} p_i s_i(\omega) - \sum_{i \in S} \sum_{w \in Z^S} f_{wi} A_{wi}(\omega) \right), \tag{5.13a}$$

$$\text{s.t. } s_i(\omega) \leq \sum_{w \in Z^S} A_{wi}(\omega), \; i \in S, \tag{5.13b}$$

$$s_i(\omega) \leq X_i(\omega), \; i \in S, \tag{5.13c}$$

$$q_w - \sum_{i \in S} A_{wi}(\omega) = 0, \; w \in Z^S, \tag{5.13d}$$

$$A_{wi}(\omega), s_i(\omega) \geq 0 \quad w \in Z^S, i \in S. \tag{5.13e}$$

Here, $s_i(\omega)$ denotes actual sales at retailer i and $A_{wi}(\omega)$ is the quantity transferred from warehouse w to retailer i. The first constraint states that total sales cannot exceed the initial allocation, while the second constraint states that sales cannot exceed demand. The last constraint implies that warehouses don't keep inventory and all ordered units should be transferred to the retailers. Thus, in the second stage of this model, we determine the allocation of ordered products that maximizes the revenue after the demand is observed, while in the first stage, we determine the order quantity, which maximizes the expected profit of the coalition. Denote by $\alpha_i(\omega), \beta_i(\omega)$, and $\gamma_w(\omega)$ ($\omega \in \Omega$, $i \in S$, $w \in Z^S$), the dual variables associated with constraints (5.13b)–(5.13d) . Then, the dual of above two-stage stochastic program can be written as

$$\min \mathbb{E} \left[\sum_{i \in S} X_i(\cdot) \beta_i(\cdot) \right] \tag{5.14}$$

$$\begin{aligned} \text{s.t. } & \mathbb{E}[\gamma_w(\cdot)] \geq -c_w, \quad w \in Z^S \\ & \alpha_i(\omega) + \beta_i(\omega) \geq p_i, \quad i \in S, \omega \in \Omega \\ & \alpha_i(\omega) + \gamma_w(\omega) \leq f_{wi}, \quad i \in S, w \in Z^S, \omega \in \Omega \\ & \alpha_i(\omega), \beta_i(\omega) \geq 0, \quad i \in S, \omega \in \Omega. \end{aligned}$$

Results of Rockafellar and Wets (1976) on strong duality of stochastic linear programs provide the following corollary.

Corollary 5.4.7 (Chen and Zhang 2009). Let Γ^w be a newsvendor situation with warehouses. Then for any coalition, $S \subseteq N$, $v(S)$ is equal to the optimal value of (5.14).

Suppose that $((\alpha_i^*(\omega))_{i \in N, \omega \in \Omega}, (\beta_i^*(\omega))_{i \in N, \omega \in \Omega}, (\gamma_w^*(\omega))_{w \in W, \omega \in \Omega})$ is an optimal solution of the dual (5.14) with $S = N$ and define

$$y_i = \mathbb{E}\left[X_i(\cdot) \beta_i^*(\cdot)\right] \text{ for all } i \in N. \tag{5.15}$$

Theorem 5.4.8 (Chen and Zhang 2009). *Let Γ^w be a newsvendor situation with warehouses. Then $\langle N, v^{\Gamma^w}\rangle$ is balanced and the vector $(y_i)_{i\in N}$, defined by (5.15), is in the core.*

The advantage of the duality approach of Chen and Zhang (2009) is that it provides a constructive proof of core nonemptiness by identifying one element in it. Using a duality approach, Chen and Zhang (2009) also derived a closed-form expression of a core element for the standard newsvendor games in Sect. 5.3. If demand has a continuous distribution, their closed-form expression results in the allocation rule in Theorem 5.3.2, which is shown by Montrucchio and Scarsini (2007) for newsvendor games with finitely and infinitely many number of players.

5.4.4 Core and Profit Allocation per Demand Realization

We remark that newsvendor games with warehouses consider the profit of all possible coalitions in expectation. Therefore, the core contains stable allocations of joint profit in expectation, which would be the main criterion of a retailer to join the grand coalition. However, for real-life application, one needs to translate a stable allocation of expected joint profit into a mechanism that allocates the realized profit at the end of the day. Özen et al. (2008) studied this issue in a noncooperative game setting in which the retailers strategically choose the coalition they want to join, the contract they want to sign to allocate joint profit, and the order vector after joining the coalition. Let therefore Γ^w be a newsvendor situation with warehouses and recall that $Q^{\{i\}}$ denotes the set of order vectors of retailer i. Let $\bar{Q}^S = \times_{i\in S} Q^{\{i\}}$ be the collection of order vectors of retailers in coalition S. Note that $\bar{Q}^S$ is different from Q^S because in this noncooperative setting every player chooses her own order vector even if she belongs to a coalition.

Consider coalition $S \subseteq N$. Let Υ^S denote the set of realizations of random vector X^S. Let $a^S : \bar{Q}^S \times \Upsilon^S \to \mathbb{R}^{m\times n}$ be a *recourse action function*, which determines an allocation of total orders for any demand realization x^S of X^S and any order profile $\bar{q}^S \in \bar{Q}^S$. We call a recourse action function a^S *feasible* if for all $\bar{q}^S \in \bar{Q}^S$ and every realization x^S of X^S,

$$a^S(\bar{q}^S, x^S) \in M^S\left(\sum_{i\in S} \bar{q}_i^S\right).$$

The set of all feasible recourse action functions for coalition S is denoted by $\mathscr{A}^S$.

Let $\bar{\pi}^S : \left\{(\bar{q}, m, x^S) \;\middle|\; \bar{q} \in \bar{Q}^S,\ x^S \in \Upsilon^S,\ m \in M^S(\sum_{i\in S} \bar{q}_i)\right\} \to \mathbb{R}^S$ be a *monetary transaction function* that allocates the profit made by the coalition S among its members. We call a monetary transaction function $\bar{\pi}^S$ *efficient*, if for all $\bar{q}^S \in \bar{Q}^S$, every realization x^S of X^S and every recourse action $m^S \in M^S(\sum_{i\in S} \bar{q}_i^S)$, we have

$$\sum_{i\in S} \bar{\pi}_i^S(\bar{q}^S, m^S, x^S) = -C\left(\sum_{i\in S} \bar{q}_i^S\right) + R^S(m^S, x^S).$$

The set of all efficient monetary transaction functions of coalition $S \subseteq N$ is denoted by $\mathscr{E}^S$.

Definition 5.4.9. An *admissible contract* for coalition $S \subseteq N$ is a pair $(a^S, \bar{\pi}^S)$ consisting of a feasible recourse action function $a^S \in \mathscr{A}^S$ and an efficient monetary function $\bar{\pi}^S \in \mathscr{E}^S$. The set of all admissible contracts for coalition $S \subseteq N$, is denoted by $\mathfrak{C}^S$, i.e., $\mathfrak{C}^S := \{(a,\bar{\pi}) | a \in \mathscr{A}^S; \bar{\pi} \in \mathscr{E}^S\}$.

Definition 5.4.10. A *profit-sharing contract* for coalition $S \subseteq N$ is a pair $(a^S, \bar{\pi}^S) \in \mathfrak{C}^S$ such that, for all $\bar{q}^S \in \bar{Q}^S$ and every realization x^S of X^S,

$$a^S(\bar{q}^S, x^S) \in \underset{m \in M^S(\sum_{i \in S} \bar{q}_i^S, x^S)}{\arg\max} \left\{ R^S(m, x^S) \right\}$$

and there exist $\lambda_i > 0$ for all $i \in S$ such that $\sum_{i \in S} \lambda_i = 1$ and for all $\bar{q}^S \in \bar{Q}^S$, x^S of X^S and $m^S \in M^S(\sum_{i \in S} \bar{q}_i^S, x^S)$

$$\bar{\pi}_i^S(\bar{q}^S, m^S, x^S) = \lambda_i \left(-C(\sum_{i \in S} \bar{q}_i^S) + R^S(m^S, x^S) \right).$$

Briefly, profit-sharing contracts are admissible contracts that choose the optimal allocation of total orders for each demand realization and divide the total profit proportionally according to preset rates among the coalition members. Under profit-sharing contracts, it is all players' interest to increase the total profit as they get a fixed percentage of it, and hence these contracts induce strategies that maximize total expected profit by the players. The set of all profit-sharing contracts for coalition $S \subseteq N$ is denoted by $\mathfrak{P}^S$.

Consider a coalition $S \subseteq N$ and an admissible contract $c^S = (a^S, \bar{\pi}^S) \in \mathfrak{C}^S$. Then, the expected payoff of player $i \in S$ in contract c^S for order profile $\bar{q}^S \in \bar{Q}^S$ is denoted by

$$\Pi_i^{c^S}(\bar{q}^S) = \mathbb{E}[\bar{\pi}_i^S(\bar{q}^S, a^S(\bar{q}^S, X^S), X^S)].$$

Now, we can define the associated noncooperative game by $\Lambda = \langle N, (T_i)_{i \in N}, (K_i)_{i \in N} \rangle$, where T_i and $K_i : \times_{i \in N} T_i \to \mathbb{R}$ represent the extended strategy space and payoff function of retailer $i \in N$, respectively. For each $i \in N$, the *extended strategy space* T_i is defined by

$$T_i := \{(S, c, q) | S \subseteq N \text{ with } i \in S, c \in \mathfrak{C}^S, \text{ and } q \in Q^i\}$$

and for every $t = (S_j, c_j, q^j)_{j \in N} \in \times_{j \in N} T_j$, the *payoff function* K_i is given by

$$K_i(t) = \begin{cases} \Pi_i^{c_i}(q^{S_i}) & \text{if } S_i = S_j \text{ and } c_i = c_j \text{ for all } j \in S_i, \\ \tau^{\{i\}}(q^i) & \text{otherwise,} \end{cases}$$

where $q^{S_i} = (q^j)_{j \in S_i}$.

In this game in strategic form, a coalition can only be formed if all the players in a contract agree upon it. If a coalition is formed, recourse actions are taken and the profit is shared as described in the contract by its recourse action function and its monetary transaction function. The players that do not form a coalition are considered as if they worked alone. The following theorem shows that the set of payoff vectors resulting from strong Nash equilibria coincides with the core.

Theorem 5.4.11 (Özen et al. 2008). *Let Γ^w be a newsvendor situation with warehouses and $\Lambda = \langle N, (T_i)_{i\in N}, (K_i)_{i\in N}\rangle$ the related game in strategic form. For all $y \in \text{core}(\langle N, v^{\Gamma^w}\rangle)$, there exists $t_N \in \mathscr{S}(\Lambda)$ such that $(K_i(t_N))_{i\in N} = y$. Moreover, $(K_i(t_N))_{i\in N} \in \text{core}(\langle N, v^{\Gamma^w}\rangle)$ for all $t_N \in \mathscr{S}(\Lambda)$.*

The proof of this theorem shows how to construct a strong Nash equilibrium that results in a specific core element using profit-sharing contracts.

Proposition 5.4.12 (Özen et al. 2008). *Let Γ^w be a newsvendor situation with warehouses and let $y \in \text{core}(\langle N, v^{\Gamma^w}\rangle)$. Consider a strategy profile $t_N = (N, p^N, q_i)_{i\in N}$, such that $p^N \in \mathfrak{P}^N$ with $\lambda_i = y_i/v(N)$ for all $i \in N$ and $\sum_{i\in N} q_i = q^{N,*}$, where $q^{N,*}$ is the optimal order vector of the grand coalition. Then, $K_i(t_N) = y_i$ for all $i \in N$ and $t_N \in \mathscr{S}(\Lambda)$.*

Therefore, the profit-sharing contract as constructed in Proposition 5.4.12 allocates the realized total profit efficiently and, in expectation, each retailer gets a stable allocation.

Hartman and Dror (2005) studied realization games defined by characteristic function $v(S,\omega) = -C(q^{*,S}) + r^S(q^{*,S}, X^S(\omega))$, for each demand scenario ω. They showed that the core of these games can be empty. Dror et al. (2008) considered a repeated cost-allocation scheme for dynamic realization games based on some rules proposed by Lehrer (2002) and they proved that the cost subsequences of the dynamic realization game process converge almost surely to the core of the expected game. This is the case even when the one period realization games have an empty core in general, as in Hartman and Dror (2005).

5.4.5 Nonlinear Costs and Revenues

The newsvendor situations with warehouses studied so far assume linear costs and revenues. Several extensions of this cost structure have been studied in the literature. Chen (2009) considered *newsvendor situations with warehouses and pricing*. Such a situation can be modeled as a tuple $\Gamma^{wp} = \langle N, W, (Z_i)_{i\in N}, (\alpha_i)_{i\in N}, (\beta_i)_{i\in N}, (k_w)_{w\in W}, (f_{wi})_{w\in W, i\in N}, (t_i)_{i\in N}, (\underline{p}_i)_{i\in N}, (\overline{p}_i)_{i\in N}\rangle$, where N, W, $(Z_i)_{i\in N}$, and $(f_{wi})_{w\in W,i\in N}$ are as before and

1. α_i and β_i are nonnegative random variables such that the demand for retailer $i \in N$ equals $X_i(\omega) = \beta_i(\omega) - \alpha_i(\omega) \cdot p_i$ for all $\omega \in \Omega$ and $p_i \in [\underline{p}_i, \overline{p}_i]$.
2. $k_w : \mathbb{R}_+ \to \mathbb{R}$ is the cost function for warehouse w.

3. t_i is the unit penalty cost for unsatisfied demand for retailer $i \in N$.
4. $\underline{p}_i$ and $\overline{p}_i$ are the lower and upper bounds for the pricing decisions of retailer $i \in N$.

Different from the multiple warehouse model in Sect. 5.4.2, retailers have to decide upon their selling prices, thereby determining their demand. Moreover, to avoid technicalities, it is assumed that Ω is finite. Chen (2009) distinguishes two types of models, the *postponed pricing model* and the *nonanticipative pricing model*. In the postponed pricing model, every retailer $i \in N$ decides upon his selling price $p_i(\omega)$ after realization of the true state of the world $\omega \in \Omega$, whereas in the nonanticipative pricing model the pricing decision is made before realization of ω.

In order to define the corresponding cooperative games let $S \subseteq N$. Let $q^S \in Q^S$ be a joint order vector, $A^S \in M^S(q^S)$ an allocation of this joint order and $p^S = (p_i(\omega))_{i\in S,\omega\in\Omega} \in \mathbb{R}_+^{S\times\Omega}$ be a vector of postponed pricing decisions. The coalition's total revenue in state of the world ω is

$$\begin{aligned} &K^S(A^S, p^S(\omega), \omega) \\ &= \sum_{i\in S}\Bigg(-\sum_{w\in Z_S} A^S_{wi} f_{wi} + p_i(\omega)\min\Bigg\{\sum_{w\in Z_S} A^S_{wi}, \beta_i(\omega) - \alpha_i(\omega)p_i(\omega)\Bigg\} \\ &\qquad - t_i \max\Bigg\{\beta_i(\omega) - \alpha_i(\omega)p_i(\omega) - \sum_{w\in Z_S} A^S_{wi}, 0\Bigg\}\Bigg), \end{aligned} \tag{5.16}$$

where $p^S(\omega) = (p_i(\omega))_{i\in S}$. As in Sect. 5.4.2 we can find, for every $\omega \in \Omega$, an allocation $A^{S,*}(\omega) \in M^S(q^S)$ that maximizes (5.16). The expected profit function of coalition S is defined as

$$\bar{\pi}^S(q^S, p^S) = \mathbb{E}\left[K^S(A^{S,*}(\cdot), p^S(\cdot), \cdot)\right] - \sum_{w\in W} k_w(q^S_w).$$

The *general newsvendor game with warehouses and postponed pricing* is given by the characteristic function $v_1^{\Gamma^{wp}} : 2^N \to \mathbb{R}$, defined by

$$v_1^{\Gamma^{wp}}(S) = \max_{q^S\in Q^S, p^S\in\mathbb{R}_+^{S\times\Omega}} \bar{\pi}^S(q^S, p^S) \text{ for all } S \subseteq N.$$

In the nonanticipative pricing model the corresponding cooperative game $v_2^{\Gamma^{wp}}$ is defined in a completely analogous way, the only difference being the fact that the selected prices do not depend upon ω, i.e., $p^S = (p_i)_{i\in S} \in \mathbb{R}^S$.

The first result considers linear cost functions.

Theorem 5.4.13 (Chen 2009). *Let* $\Gamma^{wp} = \langle N, W, (Z_i)_{i\in N}, (\alpha_i)_{i\in N}, (\beta_i)_{i\in N}, (k_w)_{w\in W}, (f_{wi})_{w\in W, i\in N}, (t_i)_{i\in N}, (\underline{p}_i)_{i\in N}, (\overline{p}_i)_{i\in N}\rangle$ *be a newsvendor situation with warehouses and pricing, such that the cost functions are linear, i.e., for every* $w \in W$ *there is a* *constant c_w such that $k_w(q) = c_w q$ for every $q \in \mathbb{R}_+$. Then $\langle N, v_1^{\Gamma^{wp}} \rangle$ and $\langle N, v_2^{\Gamma^{wp}} \rangle$ both have a nonempty core.*

The following theorem considers newsvendor situations with one warehouse and nonlinear cost functions.

Theorem 5.4.14 (Chen 2009). *Let $\Gamma^{wp} = \langle N, W, (Z_i)_{i\in N}, (\alpha_i)_{i\in N}, (\beta_i)_{i\in N}, (k_w)_{w\in W}, (f_{wi})_{w\in W, i\in N}, (t_i)_{i\in N}, (\underline{p}_i)_{i\in N}, (\overline{p}_i)_{i\in N} \rangle$ be a newsvendor situation with warehouses and pricing, such that $|W| = 1$ and such that, for every $w \in W$, the cost function k_w satisfies the following three properties:*

1. *$q \mapsto k_w(q)/q$ is nonincreasing*
2. *k_w is lower semicontinuous*
3. *$k_w(q) \to \infty$ if $q \to \infty$*

Then $\langle N, v_1^{\Gamma^{wp}} \rangle$ has a nonempty core. If moreover there exists constants $f_w, w \in W$, such that $f_{wi} = f_w$ for every $w \in W$ and $i \in N$, and a constant t such that $t_i = t$ for every $i \in N$, i.e., the transportation and penalty costs do not depend upon the retailers, then $\langle N, v_2^{\Gamma^{wp}} \rangle$ has a nonempty core as well.

We remark that this theorem considers situations with a single warehouse and symmetric revenue functions. This result does not hold for games with multiple warehouses even under concave cost functions. In a counterexample, Chen and Zhang (2009) showed that the core of these games with concave ordering cost can be empty. On the other hand, they still managed to show that the core of the game is nonempty if the solution of the dual problem satisfies certain sufficient conditions.

A second generalization of newsvendor games with warehouses considers the revenue side. Özen et al. (2009) considered a general framework to study cooperation under uncertainty. This framework covers the *general newsvendor situations with warehouses*. Such a situation is a tuple $\Gamma^{gw} = \langle N, W, (Z_i)_{i\in N}, (X_i)_{i\in N}, k, (H_i)_{i\in N} \rangle$ where N, W, $(Z_i)_{i\in N}$, and $(X_i)_{i\in N}$ are as before and

1. $k : \mathbb{R}^W \to \mathbb{R}$ is a positively homogeneous convex function.
2. $H_i : \mathbb{R}^W \times \Upsilon^{\{i\}} \to \mathbb{R}$ is for every $i \in N$ a function such that $H_i(\cdot, x_i)$ is concave for every $x_i \in \Upsilon^{\{i\}}$. As before, the set $\Upsilon^{\{i\}}$ denotes the set of realizations of random variable X_i.

In order to define the corresponding cooperative game again let $S \subseteq N$, let $q^S \in Q^S$ be a joint order vector, and let $A^S \in M^S(q^S)$ be an allocation of this joint order. The coalition's total revenue in state of the world ω is

$$R^S(A^S, \omega) = \sum_{i\in S} H_i((A^S_{wi})_{w\in W}, X_i(\omega)). \tag{5.17}$$

For every $\omega \in \Omega$, let allocation $A^{S,*}(\omega) \in M^S(q^S)$ be a maximizer of (5.17). The expected profit function of coalition S is defined as

$$\tilde{\pi}^S(q^S) = \mathbb{E}\left[R^S(A^{S,*}(\cdot), \cdot)\right] - k(q^S).$$

The *general newsvendor game with warehouses* is given by the characteristic function $v^{\Gamma^{gw}} : 2^N \to \mathbb{R}$, defined by

$$v^{\Gamma^{gw}}(S) = \max_{q^S \in Q^S} \tilde{\pi}^S(q^S) \text{ for all } S \subseteq N.$$

Theorem 5.4.15 (Özen et al. 2009). *Every general newsvendor game with warehouses has a nonempty core.*

5.4.6 Other Results on the Core

Several other aspects of cooperative newsvendor games are studied in the literature. For instance, Hartman and Dror (2003) studied the cost game among the retailers with normally distributed and correlated individual demands. In their game, the value of a coalition is a function of the covariance matrix and each coalition can manipulate the correlations to minimize costs. They use a greedy approach and the nucleolus, which is a one-point solution concept for TU games introduced by Schmeidler (1969) that always selects a core element whenever one exists, as solution. Burer and Dror (2011) extended their analysis and provided a closed form solution for this optimization problem and the nucleolus of the game.

Hartman and Dror (2005) studied newsvendor cost games with nonidentical holding and penalty costs. After showing that the core of these games might be empty, they derived the conditions under which such a game will be subadditive.

Özen et al. (2012a) studied newsvendor games with delivery restrictions. In their model, every retailer poses a constraint on the allocation of the joint orders to guarantee supply if their realized demand is high. After showing that the core of these games is nonempty, they investigated whether the core of these games satisfies certain monotonicity properties under dynamic system parameters.

Özen et al. (2012b) considered an extension of general newsvendor games, in which the allocation of the joint orders takes place after the retailers receive a demand signal and update their forecasts (not necessarily after demand realization). They studied two types of cooperation. In the first type of cooperation, the retailers allocate the joint order after sharing their updated forecasts. In a counterexample, it has been shown that collaboration with forecast sharing can harm the coalition if the retailers have asymmetric forecasting capabilities. In these cases, stability of the grand coalition is not guaranteed. However, if the retailers possess symmetric forecasting capabilities, the core of the associated game is nonempty. In the second type of cooperation, the retailers enroll in joint forecasting activities by sharing market information and the joint order is allocated using the joint forecasts. For the related games, the core is nonempty.

5.4.7 Other Approaches and Related Models

Finally, we review some papers studying other (partial) cooperative models of decentralized distribution systems.

Anupindi et al. (2001) studied a two-stage model where in the first stage the retailers decide on the order quantities noncooperatively, whereas, in the second stage, they cooperate by sharing their excess demand or supply through lateral transshipment. They showed that the second stage cooperative game has a nonempty core and a core allocation can be derived using dual prices of the transshipment problem. Moreover, they showed that a profit-allocation mechanism based on these dual prices can coordinate the system by inducing the retailers to order system optimal quantities.

Anupindi et al. (2001) assumed that retailers share all of their available excess demand or supply with other retailers in the second stage. Granot and Sošić (2003) relaxed this assumption by introducing an intermediate stage to the model, in which retailers decide noncooperatively on how much of their excess demand or supply to share with other retailers. They showed that a core allocation mechanism may not induce the retailers to share their entire excess amounts, which reduces system profit. On the other hand, the Shapley value and the fractional rule have this property but they are not necessarily in the core.

Dror and Hartman (2007) studied cost allocation in a multiple product inventory system, when consolidation of shipments is possible. They constructed a cooperative game and showed conditions for the core to be nonempty and then examined the sensitivity of these conditions to the parameters of the model.

Recently another stability concept, namely farsighted stability (Chwe 1994), has received attention in several papers. Different from the core concept, which only considers immediate deviations from the grand coalition, farsighted stability considers further deviations that the players can take as a reaction to a previous deviation. Hence, a farsighted player would consider these further deviations that can take place before deciding to deviate from the grand coalition. If the outcome of these series of deviations is not beneficial, the player would not deviate in the first place even if it improves his position temporarily, i.e., in myopic sense. Building on the results of Anupindi et al. (2001) and Granot and Sošić (2003), Sošić (2006) studied the Shapley value, which does not necessarily result in core allocations. Sošić (2006) showed that the Shapley value allocations are stable in farsighted sense. Kemahlioğlu-Ziya and Bartholdi (2011) considered a group of retailers who share a common supplier who keeps separate stock for each retailer and bears all of the inventory risk. They investigated the cooperative game, in which the players can form inventory-pooling coalitions instead of keeping separate stocks and hence increase total profit. They proposed a mechanism based on the Shapley value and showed that its allocation is stable in farsighted sense. Finally, they focus on the retailers' collusion against the supplier, which, contrary to intuition, turns out to be not always profitable for the retailers.

Hanany and Gerchak (2008) studied newsvendor cooperation in a non-TU (nontransferable utility) game framework. In their model, the retailers negotiate on the inventories they want to keep and their entitlements on these inventories in a shortfall situation. Therefore, they cooperate only exchanging inventories but not money. They study the outcome of this cooperation under the Nash bargaining solution and for different risk profiles of the retailer, i.e., risk neutral and risk averse.

Techniques similar to the ones described in this survey have been applied by Anily and Haviv (2010) to a queueing model where servers may improve the efficiency of the system by pooling their service capacities.

5.5 Conclusion and Future Research

In this survey, we reviewed the papers studying inventory cooperation using newsvendor models and focusing on the core concept. The fundamental question is whether the anticipated benefits can allow a stable collaboration to be formed. This question is answered affirmatively by studying the core of the associated games for the basic newsvendor models and the extensions we discussed here. However, there are several questions that are open for further study.

Although it has been shown (by either identifying a core element or showing that the game is balanced) that the core of cooperative newsvendor games is nonempty, not much has been done on how one of these core allocations should or would be selected. It would make an interesting research direction to investigate the characteristics (e.g., computational complexity or other fairness arguments that the retailers might ask for) that the core elements or other know solution concepts (e.g., the nucleolus and the Shapley value) satisfy.

In the analysis of retailers' cooperation, it is assumed that there exist binding agreements between the retailers once they form a coalition (which is a main assumption in cooperative game theory). However, this type of collaboration would involve coordination of participants' actions and sharing of private information in a cooperative fashion. Although economical benefits might be enough to motivate the retailers to participate in these actions, there might be other factors or mechanisms that might lead to this type of collaboration. One such example is given by Özen et al. (2008), which is also reviewed in Sect. 5.4.4. They showed that a profit-sharing contract with the rates determined by a core allocation can induce the retailers to coordinate their actions and form the grand coalition. There might be other classes of contracts with similar properties. Moreover, proper penalty and review strategies can be devised as the retailers are in a committed long-term cooperation (see e.g., Ren et al. 2010). Finally, there might be other sociological mechanisms in effect. For example, even in the absence of any incentive compatible mechanism, it has also been argued in the literature (see Maskin and Riley 1984; Crawford and Sobel 1982; Crawford 1998) that effective information sharing and coordination can be achieved through cheap talk if the parties share enough common interests. In another example paper, Özer et al. (2011) studied the role of trust in supply chain cooperation. Detailed modeling of retailers' interactions and further analysis of incentive mechanisms in this cooperative context would make an interesting research direction.

References

Anily, S., & Haviv, M. (2010). Cooperation in service systems. *Operations Research*, *58*(3): 660–673.

Anupindi, R., & Bassok, Y. (1999). Centralization of stocks: retailers vs. manufacturer. *Management Science*, *45*(2), 178–191.

Anupindi, R., Bassok, Y., & Zemel, E. (2001). A general framework for the study of decentralized distribution systems. *Manufacturing & Service Operations Management*, *3*(4), 349–368.

Aumann, R. J. (1959). Acceptable points in general cooperative *n*-person games. In A. Tucker & R. Luce (Eds.), *Contributions to the theory of games IV* (pp. 287–324). Princeton: Princeton University Press.

Bondareva, O. N. (1963). Some applications of the methods of linear programming to the theory of cooperative games. *Problemy Kibernet.*, *10*, 119–139.

Burer, S., & Dror, M. (2011). Newsvendor games: convex optimization of centralized inventory operations. *TOP*. forthcoming.

Cachon, G., & Netessine, S. (2004). Game theory in supply chain analysis. In D. Simchi-Levi, S. D. Wu, & M. Shen (Eds.), *Handbook of quantitative supply chain analysis: Modeling in the e-business era*. Berlin: Springer.

Chen, X. (2009). Inventory centralization games with price-dependent demand and quantity discount. *Operations Research*, *57*(6), 1394–1406.

Chen, X., & Zhang, J. (2009). A stochastic programming duality approach to inventory centralization games. *Operations Research*, *57*(4), 840–851.

Chwe, M. S.-Y. (1994). Farsighted coalitional stability. *Journal of Economics Theory*, *63*, 299–325.

Crawford, V. (1998). A survey of experiments on communication via cheap talk. *Journal of Economics Theory*, *78*, 286–298.

Crawford, V., & Sobel, J. (1982). Strategic information transmission. *Econometrica*, *50*, 1431–1451.

Dror, M., Guardiola, L. A., Meca, A., & Puerto, J. (2008). Dynamic realization games in newsvendor inventory centralization. *International Journal of Game Theory*, *37*(1), 139–153.

Dror, M., & Hartman, B. C. (2007). Shipment consolidation: who pays for it and how much? *Management Science*, *53*(1), 78–87.

Dror, M., & Hartman, B. C. (2011). Survey of cooperative inventory games and extensions. *Journal of the Operations Research Society*, *62*(4), 565–580.

Eppen, G. D. (1979). Effect of centralization on expected cost in a multi-location newsboy problem. *Management Science*, *25*, 498–501.

Fiestras-Janeiro, M., García-Jurado, I., Meca, A., & Mosquera, M. (2011). Cooperative game theory and inventory management. *European Journal Operations Research*, *210*, 459–466.

Gerchak, Y., & Gupta, D. (1991). On apportioning costs to customers in centralized continuous review inventory systems. *Journal of Operations Management*, *10*, 546–551.

Granot, D., & Sošić, G. (2003). A three-stage model for a decentralized distribution system of retailers. *Operations Research*, *51*(5), 771–784.

Hanany, E., & Gerchak, Y. (2008). Nash bargaining over allocations in inventory pooling contracts. *Naval Research Logistics*, *55*(6), 541–550.

Hartman, B. C. (1994). *Cooperative games and inventory cost allocation*. PhD thesis, University of Arizona.

Hartman, B. C., & Dror, M. (1996). Cost allocation in continuous review inventory models. *Naval Research Logistics*, *43*(1), 549–561.
Hartman, B. C., & Dror, M. (2003). Optimizing centralized inventory operations in a cooperative game theory setting. *IIE Transactions*, *35*(3), 243–257.
Hartman, B. C., & Dror, M. (2005). Allocation of gains from inventory centralization in newsvendor environments. *IIE Transactions*, *37*(2), 93–107.
Hartman, B. C., Dror, M., & Shaked, M. (2000). Cores of inventory centralization games. *Games and Economic Behavior*, *31*(1), 26–49.
Ichiishi, T. (1990). Comparative cooperative game theory. *International Journal of Game Theory*, *19*(2), 139–152.
Kannai, Y. (1969). Countably additive measures in cores of games. *Journal of Mathematical Analysis and Applications*, *27*, 227–240.
Kemahlioğlu-Ziya, E., & Bartholdi, J. (2011). Centralizing inventory in supply chains by using shapley value to allocate the profits. *Manufacturing & Service Operations Management*, *13*, 146–162.
Lehrer, E. (2002). Allocation processes in cooperative games. *International Journal of Game Theory*, *31*(3), 341–351.
Leng, M., & Parlar, M. (2005). Game theoretic applications in supply chain management: a review. *INFOR Information Systems and Operational Research*, *43*(3), 187–220.
Marinacci, M., & Montrucchio, L. (2004). Introduction to the mathematics of ambiguity. In I. Gilboa (Ed.), *Uncertainty in economic theory*, (pp. 46–107). London: Routledge.
Maskin, E., & Riley, M. (1984). Monopoly with incomplete information. *Rand Journal of Economics*, *15*, 171–196.
Montrucchio, L., & Scarsini, M. (2007). Large newsvendor games. *Games and Economic Behavior*, *58*(2), 316–337.
Müller, A., Scarsini, M., & Shaked, M. (2002). The newsvendor game has a nonempty core. *Games and Economic Behavior*, *38*(1), 118–126.
Nagarajan, M., & Sošić, G. (2008). Game-theoretic analysis of cooperation among supply chain agents: review and extensions. *European Journal Operations Research*, *187*(3), 719–745.
Özen, U., Erkip, N., & Slikker, M. (2012a). Stability and monotonicity in newsvendor situations. *European Journal Operations Research*, *218*(2), 416-425.
Özen, U., Fransoo, J., Norde, H., & Slikker, M. (2008). Cooperation between multiple newsvendors with warehouses. *Manufacturing & Service Operations Management*, *10*(2), 311–324.
Özen, U., Norde, H., & Slikker, M. (2011). On the convexity of newsvendor games. *International Journal of Production Economics*, *133*(1), 35–42.
Özen, U., Slikker, M., & Norde, H. (2009). A general framework for cooperation under uncertainty. *Operations Research Letters*, *37*(3), 148–154.
Özen, U., Sošić, G., & Slikker, M. (2012b). A multi-retailer decentralized distribution system with updated demand information. *European Journal Operations Research*, *216*, 573–583.
Özer, O., Zheng, Y., & Chen, K. (2011). Trust in forecast information sharing. *Management Science*, *57*, 1111–1137.
Parlar, M. (1988). Game theoretic analysis of the substitutable product inventory problem with random demands. *Naval Research Logistics*, *35*(3), 397–409.
Peleg, B., & Sudhölter, P. (2007). *Introduction to the theory of cooperative games* (2nd ed.). Berlin: Springer.
Puccetti, G., & Scarsini, M. (2010). Multivariate comonotonicity. *Journal of Multivariate Analysis*, *101*(1), 291–304.
Ren, Z., Cohen, M., Ho, T., & Terwiesch, C. (2010). Information sharing in a long-term supply chain relationship: the role of customer review strategy. *Operations Research*, *58*, 81–93.
Robinson, L. W. (1993). A comment on Gerchak and Gupta's "on apportioning costs to customers in centralized continuous review inventory systems". *Journal of Operations Management*, *11*, 99–102.
Rockafellar, R. T., & Wets, R. J.-B. (1976). Stochastic convex programming: relatively complete recourse and induced feasibility. *SIAM Journal Control Optimization*, *14*(3), 574–589.

Schmeidler, D. (1967). On balanced games with infinitely many players. Technical report, Department of Mathematics, The Hebrew University of Jerusalem. Research program in game theory and mathematical economics, research memorandum no. 28.

Schmeidler, D. (1969). The nucleolus of a characteristic function game. *SIAM Journal on Applied Mathematics*, *17*, 1163–1170.

Schmeidler, D. (1972). Cores of exact games. I. *Journal of Mathematical Analysis and Applications*, *40*, 214–225.

Shapley, L. S. (1953). A value for *n*-person games. In *Contributions to the theory of games, vol. 2*, Annals of mathematics studies (vol. 28, pp. 307–317). Princeton, NJ: Princeton University Press.

Shapley, L. S. (1967). On balanced sets and cores. *Naval Research Logistics Quarterly*, *14*, 453–460.

Shapley, L. S. (1971). Cores of convex games. *International Journal of Game Theory*, *1*, 11–26.

Slikker, M., Fransoo, J., & Wouters, M. (2001). Joint ordering in multiple news-vendor situations: a game-theoretical approach. Technical report, Eindhoven University of Technology, Eindhoven, The Netherlands.

Slikker, M., Fransoo, J., & Wouters, M. (2005). Cooperation between multiple news-vendors with transshipments. *European Journal Operations Research*, *167*(2), 370–380.

Sošić, G. (2006). Transshipment of inventories among retailers: myopic versus farsighted stability. *Management Science*, *52*, 1493–1508.

Wang, Q., & Parlar, M. (1994). A three-person game theory model of the substitutable product inventory problem with random demands. *European Journal Operations Research*, *76*, 83–97.

Chapter 6
Inventories and Stock-out Costs in the Price-Setting Newsvendor: An Economic Interpretation

Miguel Ampudia and Michael A. Salinger

Abstract According to the Lerner rule, a firm's profit-maximizing price under certainty can be characterized with just two parameters: marginal cost and the elasticity of demand. Salinger and Ampudia (Salinger, M. A., & Ampudia, M. (2011). Simple economics of the price-setting newsvendor problem. Managament Science, 57, 1996–1998.) showed that in the most basic version of the price-setting newsvendor (i.e., with no inventories or stock-out costs), the Lerner rule applies with suitable modifications to the definition of marginal cost and the elasticity of demand. This chapter extends that result to the more general version of the price-setting newsvendor problem that allows for stock-out costs and for unsold output to have some residual value as inventory. This extension suggests that the Lerner rule characterization can be a unifying framework for a wide variety of extensions to the price-setting newsvendor problem.

Keywords Newsvendor problem • Demand uncertainty • Mark-ups • Inventories • Lerner relationship

6.1 Introduction

In the simplest version of the newsvendor problem, a firm must choose its price and output while facing uncertain demand. Output is completely perishable, so the firm cannot carry forward any excess product to be sold in future periods, and

M. Ampudia
Department of Economics, Boston University, 580 Commonwealth Ave.
Unit 105, Boston, MA 02215, USA

M.A. Salinger (✉)
Boston University, School of Management, 595 Commonwealth Ave.,
Boston, MA 02215, USA
e-mail: salinger@bu.edu

T.-M. Choi (ed.), *Handbook of Newsvendor Problems*, International Series in Operations Research & Management Science 176, DOI 10.1007/978-1-4614-3600-3_6,

unsatisfied demand imposes no cost other than the opportunity cost of a foregone profit opportunity. This version of the problem—henceforth, the "simple version"—can bring essential features of the solution into sharp relief. But, as is clear from the vast literature on extensions to the problem,[1] the need to commit to both price and output in the face of demand uncertainty is present in a far wider set of circumstances than the assumptions underlying the simple version.

A striking feature of the existing literature on the newsvendor problem is that while it contains technically correct solutions for a wide range of extensions, it has lacked a simple characterization of the solutions. Recently, however, Salinger and Ampudia (2011) have provided such a characterization for the simple version by generalizing the Lerner rule under certainty (Lerner 1934) to allow for uncertain demand. Under certainty, a firm's profit-maximizing price depends on just two factors, marginal cost and the elasticity of demand. Salinger and Ampudia show that the same formula applies in the simple version with suitable generalizations of the relevant elasticity and marginal cost. Specifically, they show that the relevant elasticity is the elasticity of the average quantity sold[2] with respect to price and that the relevant marginal cost is the expected marginal cost of an expected unit sold (as distinct from the marginal cost of a unit produced).[3]

In this chapter, we show that the generalized Lerner relationship can serve as a unifying principle for characterizing and understanding how generalizations to the simple version affect the solution. Specifically, we derive the generalized Lerner relationship that accounts for two of the most thoroughly analyzed extensions of the problem. We allow for the possibilities that unsold output retains some residual value (presumably as inventories) and that the firm incurs a cost of unsold output in addition to the opportunity cost of the lost sale.[4] These "stock-out costs" reflect a loss in reputation that presumably affects future demand.[5] Inventory and stock-out costs affect the solution through the expected marginal cost of an expected unit sold.

---

[1]See Porteus 1990, Petruzzi and Dada 1999, 2009, and Khouja, 1999, for excellent reviews on these and other variants of the newsvendor problem.

[2]As we discuss in more detail below, the elasticity of the average quantity sold with respect to price reflects a weighted average over demand states, with the probabilty of selling all output being the weight for the demand state in which consumers demand exactly the quantity produced.

[3]As is discussed in more detail below, both factors reflect specific assumptions about how price changes in conjunction with changes in output.

[4]There are other possible extensions. For example, we maintain the assumption that the newsvendor is risk neutral and therefore seeks to maximize expected profits. Other possible extensions include allowing for an objective function other than expected profit maximization (Arcelus et al. 2012; Wang et al. 2009; Choi and Chiu 2010; Yang et al. 2011) and uncertainty in costs (Tang et al. 2011).

[5]They could also reflect direct costs to the extent that the firm has an obligation to supply. For example, a store that advertises a good at particular price might have an obligation to satisfy demand in some way; and doing so might prove costly. Stock-out costs are likely to be particularly important in extending newsvendor analysis to oligopolistic industries. See Krishnan and Winter 2007. For an empirical estimate of stock-out costs, see Matsa 2011.

They do not affect the relevant elasticity (and, therefore, the mark-up over marginal cost).

The result might be surprising at first. To the extent that uncertainty in demand affects the optimal price, it necessarily affects either marginal cost or the mark-up over marginal cost. Inventory costs and stock-out costs are associated with output the firm does not sell. One normally associates marginal cost with the cost of output the firm does sell. Thus, one might expect the effect of unsold output to enter the price through the mark-up. But the Lerner rule under uncertainty reveals the sense in which the cost of unsold output can be part of the marginal cost that enters the optimal pricing formula.

One of the long-standing puzzles in the literature on the price-setting newsvendor is the different qualitative effects of additive and multiplicative uncertainty on the optimal price (Mills 1959; Karlin and Carr 1962). Salinger and Ampudia resolve this puzzle for the simple version by showing how the form of uncertainty affects the two factors needed to characterize the optimal price. Holding the probability of satisfying all demand constant, additive uncertainty is fixed with respect to which price–output combination the firm chooses. As a result, additive uncertainty does not affect the marginal cost of an expected unit sold. It does, however, increase the elasticity of the average quantity sold with respect to price, so it lowers the optimal mark-up over marginal cost. In contrast, multiplicative uncertainty increases the marginal cost of an expected unit sold but does not affect the elasticity of the average quantity sold with respect to price or the mark-up. As we show in this chapter, the resolution of the uncertainty puzzle that the generalized Lerner rule reveals for the simple version holds for the extensions considered here. This further demonstrates the power of the generalized Lerner rule to serve as a unifying principle behind solutions to different versions of the problem.

6.2 The Model

In the price-setting newsvendor problem with inventories and stock-out costs, a firm must choose both its price and output before observing random demand. The firm incurs a constant unit cost, c, for each unit it produces. Each unit it produces but does not sell has a salvage value, v. In the simple version, $v = 0$. If $v = c$, then the firm suffers no penalty from producing earlier than necessary (in which case, the problem is no longer within the newsvendor category). A disposal cost implies $v < 0$. The firm also incurs a stock-out or shortage cost, c_s, for each unit demanded that it cannot supply.

6.2.1 The Quality Transformation

The standard approach to solving the price-setting newsvendor problem is to make the quantity produced and the price the two choice variables. While valid, this approach is not the most useful for assessing the effect of the elasticity of demand on the solution. Any measure of the sensitivity of demand to price must hold the state of demand constant.[6] In the standard formulation, the derivative of the quantity sold with respect to price does not hold the state of demand constant and therefore does not reflect just the sensitivity of demand to price.

Following Salinger and Ampudia, we instead use a "quality factor transformation," which is a generalization of the "stocking factor transformation." In the price-setting newsvendor literature, the "stocking factor" refers to an amount (either absolute in the case of additive uncertainty or a percentage in the case of multiplicative uncertainty) by which production exceeds "deterministic demand."[7] A stocking factor approach to the solution is to substitute the stocking factor for the quantity produced as a choice variable. Given additive (multiplicative) uncertainty, the additive (multiplicative) stocking factor determines the probability of being able to satisfy all demand, which is a dimension of quality. Salinger and Ampudia generalize the approach by making the probability of satisfying all demand a choice variable.[8] That approach is equivalent to the stocking factor approach for the cases of additive or multiplicative uncertainty, but does not require any specific assumption about the form of uncertainty.[9]

To implement this approach, let $Q(\varepsilon,p)$ be the inverse cumulative distribution function for the quantity demanded conditional on p, where p is price and ε is a random variable uniformly distributed between 0 and 1 with $\partial Q/\partial\varepsilon > 0$ and $\partial Q/\partial p < 0$.[10] Let x be output. The firm sells x if $Q(\varepsilon,p) \geq x$ and $Q(\varepsilon,p)$ otherwise.

Let ε^* be the probability that x is sufficient to satisfy all demand. The equation that implicitly defines ε^* is:

$$x = Q(\varepsilon^*, p). \tag{6.1}$$

Using (6.1) to substitute for x and defining the expected quantity the firm sells, $\bar{Q}(\varepsilon^*, p)$, and average demand, $\bar{D}(p)$, as:

[6]"Constant" can refer to a single state of demand or an average across a distribution of states.

[7]See Ernst 1970, Thowsen 1975, Petruzzi and Dada 1999, Petruzzi et al. 2009.

[8]Salinger and Ampudia refer to this as a "general stocking factor approach." Because the stocking factor typically refers to a physical quantity, "quality transformation" is a better characterization of the approach.

[9]See also Raz and Porteus (2006) fractile approach, which is a discrete approximation to the generalized stocking factor approach.

[10]The assumption that ε is uniformly distributed is a consequence of $Q(\varepsilon, p)$ being an inverse cumulative distribution function. It does not impose any particular form of the distribution function itself.

$$\bar{Q}(\varepsilon^*, p) = Q(\varepsilon^*, p)(1-\varepsilon^*) + \int_0^{\varepsilon^*} Q(\varepsilon; p)\,\mathrm{d}\varepsilon \tag{6.2}$$

$$\bar{D}(p) = \int_0^1 Q(\varepsilon, p)\mathrm{d}\varepsilon. \tag{6.3}$$

The expected profit function is:

$$E[\pi] = (p-c)\bar{Q}(\varepsilon^*, p) - (c-v)[Q(\varepsilon^*, p) - \bar{Q}(\varepsilon^*, p)] - c_\mathrm{s}[\bar{D}(p) - \bar{Q}(\varepsilon^*, p)]. \tag{6.4}$$

In (6.4), the first term is expected sales multiplied by the margin between price and the direct cost of producing the units sold. It resembles the profit function under certainty except that the expected quantity sold substitutes for the deterministic quantity sold (and produced). The second term, the expected cost of unsold output, is the expected quantity of unsold output multiplied by the cost per unit of unsold output. The third term is the expected cost of unsatisfied demand, estimated as the expected quantity of unsatisfied demand multiplied by the cost per unit of unsatisfied demand.

6.2.2 First-Order Conditions

The first order conditions for profit maximization are:

$$\frac{\partial E[\pi]}{\partial \varepsilon^*} = (p-c)\frac{\partial \overline{Q}(\varepsilon^*, p)}{\partial \varepsilon^*} - (c-v)\left[\frac{\partial Q(\varepsilon^*, p)}{\partial \varepsilon^*} - \frac{\partial \overline{Q}(\varepsilon^*, p)}{\partial \varepsilon^*}\right] + c_s \frac{\partial \overline{Q}(\varepsilon^*, p)}{\partial \varepsilon^*} = 0 \tag{6.5}$$

$$\frac{\partial E[\pi]}{\partial p} = \bar{Q}(\varepsilon^*, p) + (p-c)\frac{\partial \bar{Q}(\varepsilon^*, p)}{\partial p} - (c-v)\left[\frac{\partial Q(\varepsilon^*, p)}{\partial p} - \frac{\partial \bar{Q}(\varepsilon^*, p)}{\partial p}\right] - c_\mathrm{s}\left[\frac{d\bar{D}(p)}{dp} - \frac{\partial \bar{Q}(\varepsilon^*, p)}{\partial p}\right] = 0. \tag{6.6}$$

Because $\frac{\partial \bar{Q}(\varepsilon^*, p)}{\partial \varepsilon^*} = (1-\varepsilon^*)\frac{\partial Q(\varepsilon^*, p)}{\partial \varepsilon^*}$, (6.5) implies:

$$\varepsilon^* = \frac{p-c+c_\mathrm{s}}{p-v+c_\mathrm{s}}, \tag{6.7}$$

which is the familiar condition for the critical fractile in the classic (i.e., with fixed prices) newsvendor problem with inventory and stock-out costs.

Equation (6.6) is the basis for the generalized Lerner Index and, as a result, merits detailed examination. Because it is the partial first order condition with respect to p holding ε^* constant, the quantity produced (x) changes as p changes. More precisely, the change in output associated with a price change reflects the derivative of demand with respect to price in the state of demand associated with the critical fractile. Because it reflects the simultaneous change in price and output determined by the associated change in demand, this first order condition corresponds more nearly to the first order condition under certainty that gives rise to the Lerner relationship. However, except in the special case of additive uncertainty (which we discuss below), the effect of price on demand varies across states. Since the change in output has to be geared to a single state, the amounts of unsold output and unsatisfied demand and the associated costs vary across states.

The first two terms in (6.6) are the derivative of the first term in (6.4). The first term is the effect of a change in price on revenue holding expected quantity constant. The second term is the margin between price and direct production cost on the change in the expected quantity sold. Together, they resemble the derivative of the profit function with respect to price under certainty except that expected demand and its partial derivative with respect to price substitute for deterministic demand and its derivative.

The third term in (6.6) is the derivative of the second term in (6.4), which reflects the expected cost of unsold output. The term reflects the sense in which the cost of unsold output is part of the marginal cost of an expected unit sold. The Lerner rule under certainty is a reformulation of the firm's optimal choice of output. Under certainty, however, a change in output is necessarily accompanied by a price reduction dictated by the demand curve. Because of the quality transformation, (6.6) captures simultaneous changes in price and quantity. The rate at which output varies with price is driven by demand in the critical fractile. Because the sensitivity of demand to price can vary across states, the decision to lower price and increase output along the critical fractile can change the expected amount of unsold output. This change in unsold output associated with a change in output is a component of the marginal cost of a unit sold to be added to the direct cost of a unit sold.

The fourth term in (6.6) reflects the marginal cost of unsatisfied demand. It is a component of marginal cost for the same reason that the cost of unsold output is part of marginal cost. To the extent that the sensitivity of demand to price in the high-demand states is different than in the critical fractile, a simultaneous change in price and output along the critical fractile changes the expected amount of unsatisfied demand and must therefore be included as a marginal cost of an expected unit sold.

6.2.3 Deriving the Lerner Condition

Under certainty, the Lerner condition follows from two basic principles. First, at the profit-maximizing output, marginal revenue equals marginal cost. Second, (everywhere), marginal revenue equals $p(1 + 1/\eta)$, (where, η is the elasticity of demand). Equation (6.1) follows from equating this alternative expression for marginal revenue to marginal cost.

The same principles apply under uncertainty, with one qualification. The standard definition of marginal revenue is the additional revenue from *selling* one more unit of output (taking account of the price reduction needed to do so). The standard definition of marginal cost is the incremental revenue from *producing* an extra unit. Under certainty, the quantity produced equals the quantity sold, so distinguishing between an additional unit sold and an additional unit produced is unnecessary. But under uncertainty, the distinction matters.

As when demand is certain, at the optimal quantity produced with uncertain demand, the expected marginal revenue from an additional unit produced equals the expected marginal cost of an extra unit produced. However, since the production of an extra unit does not necessarily entail the sale of an extra unit, the expected marginal revenue from an additional unit produced has to reflect the expected increase in the quantity sold, which might be different from 1. Similarly, under uncertainty, at the optimal price and quantity produced, the marginal expected revenue from an additional expected unit sold equals the expected marginal cost of an expected unit sold. However, whenever an extra unit produced does not generate an extra expected unit sold, the additional quantity produced needed to sell an additional unit on average is different from 1. The marginal cost of an additional unit sold has to reflect the actual additional production associated with an extra expected unit sold. In summary, at the optimum under uncertainty, marginal revenue equals marginal cost on both a consistent per unit produced and a consistent per unit sold basis; but, the marginal revenue from an additional unit sold does not necessarily equal the marginal cost of an extra unit produced.

The standard definitions of marginal revenue and marginal cost are both with respect to changes in quantity. Equation (6.6) is a partial first order condition with respect to price. One can transform it into a condition on changes in quantity by dividing it by either $\partial\bar{Q}(\varepsilon^*,p)/\partial p$ or $\partial Q(\varepsilon^*,p)/\partial p$. Because the objective is to characterize the optimal price, which is revenue per unit sold, the former is more convenient:

$$\frac{\bar{Q}(\varepsilon^*,p)}{\frac{\partial\bar{Q}(\varepsilon^*,p)}{\partial p}}+p-\left\{c+\frac{(c-v)\left[\frac{\partial Q(\varepsilon^*,p)}{\partial p}-\frac{\partial\bar{Q}(\varepsilon^*,p)}{\partial p}\right]+c_{\mathrm{s}}\left[\frac{\mathrm{d}\bar{D}(p)}{\mathrm{d}p}-\frac{\partial\bar{Q}(\varepsilon^*,p)}{\partial p}\right]}{\frac{\partial\bar{Q}(\varepsilon^*,p)}{\partial p}}\right\}=0. \tag{6.8}$$

Letting $\eta_{\mathrm{A}}=\frac{\partial\bar{Q}(\varepsilon^*,p)}{\partial p}\frac{p}{\bar{Q}(\varepsilon^*,p)}$, the first two terms of (6.8) can be written as $p\left(1+{}^{1}/_{\eta_{\mathrm{A}}}\right)$, which is the expected additional revenue from an additional expected unit sold. The term in brackets is the expected marginal cost of an expected unit sold. It is the sum of two terms. The first is c, the direct marginal cost of an extra expected unit sold. The second is a fraction that reflects the marginal cost of both unsold output and unsatisfied demand. Because both marginal revenue and the direct production cost of sold output is on a per expected unit sold basis, the marginal costs of unsold output and of unsatisfied demand must also be on a per expected unit sold basis in order to be comparable. The numerator of the fraction reflects changes in the quantities produced, demanded, and sold associated with price changes. Dividing these costs by the derivative of the expected quantity sold with respect to price puts these marginal cost factors on a per expected unit sold basis.

Letting $C(\varepsilon^*, p)$ be the term in brackets in (6.8) (i.e., marginal cost of an expected unit sold given that the firm changes output with price to preserve the probability of being able to satisfy all demand), then (6.8) implies:

$$\frac{p - C(\varepsilon^*, p)}{p} = -\frac{1}{\eta_{\mathrm{A}}}. \tag{6.9}$$

Equation (6.9) is the generalization of the Lerner relationship to the price-setting newsvendor problem with inventories and stock-out costs.

6.3 Special cases

As described in the introduction, the different qualitative effects of additive and multiplicative uncertainty in the "simple price-setting newsvendor" problem arise because of their different effects on the marginal cost of an expected unit sold and the elasticity of the average quantity sold with respect to price. In this section, we demonstrate that the insight still applies with the more general version of the problem.

6.3.1 Additive Uncertainty

With additive uncertainty, $Q(\varepsilon, p) = q(p) + \Phi(\varepsilon)$, where $q(p)$ is a deterministic demand function and $\Phi(\varepsilon)$ is an inverse cumulative distribution function. As a result, $\frac{\partial Q(\varepsilon,p)}{\partial p} = \frac{\partial Q(\varepsilon^*,p)}{\partial p} \forall \varepsilon$, which implies that the term in brackets of (6.8) reduces to c, the marginal cost of producing an additional unit. The cost of unsold output remains a cost of doing business, but it is a fixed cost in that when the firm lowers price and increases output so as to hold the probability of being able to meet all demand constant, the amount (and therefore the cost associated with) unsold output does not change.

Also,

$$\eta_{\mathrm{A}} = \frac{p}{q(p) + \bar{\Phi}(\varepsilon^*)}, \tag{6.10}$$

where $\bar{\Phi}(\varepsilon^*) = \Phi(\varepsilon^*)(1 - \varepsilon^*) + \int_0^{\varepsilon^*} \Phi(\varepsilon)\mathrm{d}\varepsilon$. Thus, consider a mean-preserving spread of the distribution around ε^* so that that both low demand states and stocked-out states become more probable. Because the slope of the demand curve is constant across states, the change in the distribution of demand does not affect the slope of the average quantity sold with respect to price. It does, however, lower the average quantity sold because the additional probability of the high demand states does not generate additional sales whereas the additional probability of low demand states results in lower sales.

As a result, the elasticity of the average quantity sold with respect to price increases. In turn, the increase in the average elasticity reduces the optimal mark-up of price over marginal cost.

6.3.2 Multiplicative Uncertainty

With multiplicative uncertainty,

$$Q(\varepsilon, p) = q(p)\Psi(\varepsilon), \tag{6.11}$$

where $q(p)$ is a deterministic demand function and $\Psi(\varepsilon)$ is an inverse cumulative distribution function with $\mathrm{E}[\Psi(\varepsilon)] = 1$. Equation (6.11) implies:

$$Q(\varepsilon^*, p) = q(p)\Psi(\varepsilon^*), \tag{6.12}$$

$$\bar{Q}(\varepsilon^*, p) = q(p)\left[\Psi(\varepsilon^*)(1-\varepsilon^*) + \textstyle\int_0^{\varepsilon^*} \Psi(\varepsilon)\mathrm{d}\varepsilon\right] = q(p)\bar{\Psi}(\varepsilon^*), \tag{6.13}$$

and,

$$\frac{\partial \bar{Q}(\varepsilon^*, p)}{\partial p} = q'(p)\left[\Psi(\varepsilon^*)(1-\varepsilon^*) + \int_0^{\varepsilon^*} \Psi(\varepsilon)\mathrm{d}\varepsilon\right] = q'(p)\bar{\Psi}(\varepsilon^*). \tag{6.14}$$

Also, since $\mathrm{E}\left[\Psi(\varepsilon)\right] = 1$,

$$\bar{D}(p) = q(p). \tag{6.15}$$

Using (6.13) and (6.14), the elasticity of the average quantity sold with respect to price (η_{A}) is:

$$\eta_{\mathrm{A}} = \frac{pq'(p)\bar{\Psi}(\varepsilon^*)}{q(p)\Psi(\varepsilon^*)} = \frac{pq'(p)}{q(p)}. \tag{6.16}$$

Equation (6.16) implies that at a given price, the elasticity of the average quantity sold with respect to price is independent of the distribution of $\Psi(\varepsilon)$, which in turn implies that a mean-preserving spread of $\Psi(\varepsilon)$ around ε^* would not alter the mark-up of the optimal price over the bracketed term in (6.8), which is the marginal cost of an expected unit sold.

Turning attention to that bracketed term, (6.12)–(6.15) imply that it reduces to:

$$c + \frac{(c-v)[\Psi(\varepsilon^*) - \bar{\Psi}(\varepsilon^*)] + c_{\mathrm{s}}[1 - \bar{\Psi}(\varepsilon^*)]}{\Psi(\varepsilon^*)}. \tag{6.17}$$

Since the only effect of a mean-preserving spread around ε^* on (6.17) would be a reduction in $\bar{\Psi}(\varepsilon^*)$, increased multiplicative uncertainty raises the marginal cost of an expected unit sold by increasing both the unsold output and unmet demand components of marginal cost.

6.4 Conclusions

The generalized Lerner rule characterization of the solution to the simple version of the price-setting newsvendor extends to a more general setting and the extension demonstrates the power of this characterization to serve as a unifying principle for understanding solutions to variants of the problem. In addition to those cited in the introduction, a particularly important extension would be to place the price-setting newsvendor explicitly in a market setting.[11]

Under certainty, the Lerner rule is simply a restatement of the principle that the profit-maximizing output for a firm is where marginal revenue equals marginal cost, recognizing that the firm needs to cut prices to sell extra output. If the firm does not have to cut price at all, it gets the entire price as marginal revenue from an additional unit sold. More generally, the (perhaps negative) fraction of the price that it retains as marginal revenue depends on how much it has to cut price to stimulate demand, and that in turn depends on the elasticity of demand.

The Lerner rule generalizes to the price-setting newsvendor because much of the same logic applies, but uncertainty about demand adds a complication. The Lerner rule reflects how the firm optimally picks the feasible price–quantity combinations along a demand curve. But the increase in demand due to a price decrease varies across states, so the question arises as to how to assess the relevant price–quantity trade-off.

A key part of the solution lies with the critical fractile. Holding price constant, the critical fractile is the demand state that determines the optimal output. Thus, in the analysis of whether increasing output by a unit increases expected profits, it is useful to think of the associated price cut as being determined by the demand curve defined by the critical fractile. But, to the extent that the (absolute) sensitivity of demand to price varies across demand states, that price reduction can generate a change in demand that is different from the change in output, which in turn affects both the amount of unsold output and unsatisfied demand.

[11]While the Lerner rule under certainty and the extensions to uncertainty developed here apply to any firm provided that the demand parameters and uncertainty are understood as those facing the individual firm, explicitly modeling the link between market demand uncertainty and the uncertainty facing an individual firm would greatly extend the scope for practical application.

Appendix

Second Order Conditions and Comparative Statics

The second order conditions for a maximum are:

$$\frac{\partial^2 E[\Pi]}{\partial p^2} = 2\frac{\partial \bar{Q}(\varepsilon^*,p)}{\partial p} + p\frac{\partial^2 \bar{Q}(\varepsilon^*,p)}{\partial p^2} - c\frac{\partial^2 Q(\varepsilon^*,p)}{\partial p^2}$$
$$+v\left[\varepsilon^* \frac{\partial^2 Q(\varepsilon^*,p)}{\partial p^2} - \int_0^{\varepsilon^*} \frac{\partial^2 Q(\varepsilon,p)}{\partial p^2}\mathrm{d}\varepsilon\right]$$
$$-c_\mathrm{s}\left[\int_{\varepsilon^*}^{1} \frac{\partial^2 Q(\varepsilon;p)}{\partial p^2}\mathrm{d}\varepsilon - (1-\varepsilon^*)\frac{\partial Q^2(\varepsilon^*,p)}{\partial p^2}\right] < 0 \tag{A.1}$$

$$\frac{\partial^2 E[\Pi]}{\partial p^2}\frac{\partial^2 E[\Pi]}{\partial \varepsilon^{*2}} - \left[\frac{\partial^2 E[\Pi]}{\partial p \partial \varepsilon^*}\right]^2 > 0, \tag{A.2}$$

where:

$$\frac{\partial^2 E[\Pi]}{\partial \varepsilon^{*2}} = -(p-v+c_\mathrm{s})\frac{\partial Q(p,\varepsilon^*)}{\partial \varepsilon^*}$$

and

$$\frac{\partial^2 E[\Pi]}{\partial p \partial \varepsilon^*} = (1-\varepsilon^*)\frac{\partial Q(p,\varepsilon*)}{\partial \varepsilon^*}.$$

Let $E\pi_{..}$ denote second derivatives. The comparative statics for price with respect to c is:

$$\frac{\mathrm{d}p}{\mathrm{d}c} = -\frac{E\pi_{pc} - \dfrac{E\pi_{p\varepsilon^*}E\pi_{\varepsilon^* c}}{E\pi_{\varepsilon^*\varepsilon^*}}}{E\pi_{pp} - \dfrac{E\pi^2_{p\varepsilon^*}}{E\pi_{\varepsilon^*\varepsilon^*}}}, \tag{A.3}$$

where

$$E\pi_{pc} = \frac{\partial^2 E[\Pi]}{\partial p \partial c} = -\frac{\partial Q(\varepsilon^*,p)}{\partial p}.$$

The comparative statics with respect to v and c_s are, respectively,

$$\frac{\mathrm{d}p}{\mathrm{d}v} = -\frac{E\pi_{pv} - \dfrac{E\pi_{p\varepsilon^*}E\pi_{\varepsilon^* v}}{E\pi_{\varepsilon^*\varepsilon^*}}}{E\pi_{pp} - \dfrac{E\pi^2_{p\varepsilon^*}}{E\pi_{\varepsilon^*\varepsilon^*}}} \tag{A.4}$$

and

$$\frac{\mathrm{d}p}{\mathrm{d}c_{\mathrm{s}}} = -\frac{E\pi_{pc_{\mathrm{s}}} - \dfrac{E\pi_{p\varepsilon^*}E\pi_{\varepsilon^* cs}}{E\pi_{\varepsilon^*\varepsilon^*}}}{E\pi_{pp} - \dfrac{E\pi_{p\varepsilon^*}^2}{E\pi_{\varepsilon^*\varepsilon^*}}}, \tag{A.5}$$

where

$$E\pi_{pv} = \frac{\partial^2 E[\Pi]}{\partial p \partial v} = \varepsilon^* \frac{\partial Q(\varepsilon^*, p)}{\partial p} - \int_0^{\varepsilon^*} \frac{\partial Q(\varepsilon, p)}{\partial p} \mathrm{d}\varepsilon$$

and

$$E\pi_{pc_{\mathrm{s}}} = \frac{\partial^2 E[\Pi]}{\partial p \partial c_{\mathrm{s}}} = -\int_{\varepsilon^*}^{1} \frac{\partial Q(\varepsilon; p)}{\partial p} \mathrm{d}\varepsilon + (1-\varepsilon^*) \frac{\partial Q(\varepsilon^*, p)}{\partial p}.$$

References

Arcelus, F. J., Kumar, S., & Srinivasan, G. (2012). Risk tolerance and a retailer's pricing and ordering policies within a newsvendor framework. *Omega*, 40, 188–198.

Choi, T. M., & Chiu, C. H. (2012). Mean-downside-risk and mean-variance newsvendor models: implications for sustainable fashion retailers. *International Journal of Production Economics,* 135, 552–560.

Ernst, R. L. (1970). A linear inventory model of a monopolistic firm. *Dissertation.* Berkeley: University of California.

Karlin, S., & Carr, C. R. (1962). Prices and optimal inventory policy. In K. J. Arrow, S. Karlin, & H. Scarf (Eds.) Applied probability and management science. Stanford: Stanford University Press.

Khouja, M. (1999). The single-period (news-vendor) problem: literature review and suggestions for future research. *Omega*, 27, 537–553.

Krishnan, H., & Winter, R. A. (2007). Vertical control of price and inventory. *American Economic Review*, 97, 1840–1857.

Lerner, A. P. (1934). The concept of monopoly and the measurement of monopoly power. *The Review of Economic Studies*, 1, 157–175.

Matsa, D. A. (2011). Competition and product quality in the supermarket industry. *Quarterly Journal of Economics*, 126(3): 1539–1591.

Mills, E. S. (1959). Uncertainty and price theory. *Quarterly Journal of Economics*, 73, 116–130.

Petruzzi, N. C., Dada, M. (1999). Pricing and the newsvendor problem: a review with extensions. *Operation Research*, 47, 183–194.

Petruzzi, N. C., & Dada, M. (2009). Newsvendor models, unpublished manuscript.

Petruzzi, N. C., Wee, K. E., & Dada, M. (2009). The newsvendor model with consumer search costs. Production, *Operation and Management*, 18, 693.

Porteus, E. L. (1990). Stochastic inventory theory. In D. P. Heyman, M. J. Sobel (Eds.) Handbooks in OR & MS, v 2. North-Holland: Elsevier.

Raz, G., & Porteus, E. (2006). A fractiles perspective to the joint price/quantity newsvendor model. *Management Science*, 52, 1764–1777.

Salinger, M. A., & Ampudia, M. (2011). Simple economics of the price-setting newsvendor problem. *Managament Science*, 57, 1996–1998.

Tang, O., Musa, S.-N., & Li, J. (2011). Dynamic pricing in the newsvendor problem with yield risks. *International Journal of Production Economics*, forthcoming.

Thowsen, G. T. (1975). A dynamic, non-stationery inventory problem for a price/quantity setting firm. *Naval Research Logistics Quarterly*, 22, 461–476.

Yang, S., Shi, C. V., & Zhao, X. (2011). Optimal ordering and pricing decisions for a target oriented newsvendor. *Omega*, 39, 110–115.

Wang, C. X., Webster, S., & Suresh, N. C. (2009). Would a risk-averse newsvendor order less at a higher selling price? *European Journal of Operational Research*, 196, 544–553.

Chapter 7
Newsvendor Models with Alternative Risk Preferences Within Expected Utility Theory and Prospect Theory Frameworks

Charles X. Wang, Scott Webster, and Sidong Zhang

Abstract Newsvendor models are widely used in the literature, and usually based upon the assumption of risk neutrality. Recently there is a growing body of literature that attempts to use alternative risk preferences rather than risk neutrality to describe the newsvendor decision-making behavior. In this chapter, we provide an overview of newsvendor models with alternative risk preferences within the expected utility theory and prospect theory frameworks and identify some directions for future research.

Keywords Expected utility theory • Prospect theory • Risk aversion • Loss aversion • Newsvendor model • Behavioral operations management

7.1 Introduction

The single-period newsvendor model is one of the fundamental models in inventory management. In the classical newsvendor model setting, a newsvendor who sells a short life-cycle product with uncertain demand must decide how many products to order before the season begins. If realized demand is higher than the initial order quantity, the newsvendor will face lost sales, whereas if realized demand is lower than the initial order quantity, the newsvendor will liquidate all unsold products

C.X. Wang (✉)
School of Management, State University of New York, Buffalo, NY 14260, USA
e-mail: cxwang@buffalo.edu

S. Webster
Whitman School of Management, Syracuse University, Syracuse, NY 13210, USA
e-mail: stwebste@syr.edu

S. Zhang
Chinesisch-Deutsches Hochschulkolleg, Tongji University, Shanghai 200092, China
e-mail: sidong_zhang@tongji.edu.cn

T.-M. Choi (ed.), *Handbook of Newsvendor Problems*, International Series in Operations Research & Management Science 176, DOI 10.1007/978-1-4614-3600-3_7,

at a lower salvage value. With the current business trends of shortening product life-cycles, increasing product variety, and frequent new product introductions, the single-period newsvendor model has been applied to a large variety of products, including books, consumer electronics, food, fashion apparel, personal computers, and toys, etc.

The classical newsvendor model is based upon risk neutrality, meaning that a manager will place an order to maximize the expected profit or minimize the expected cost. However, in practice, there is much evidence that managers' inventory decisions are not always consistent with profit maximization. For example, Fisher and Raman (1996) observe that a fashion apparel manufacturer orders systematically less than the profit-maximizing order quantity; Patsuris (2001) reports that despite the bad economy in 2001, many retailers continue to order more unnecessary supply; and Schweitzer and Cachon (2000) have designed two experiments to test the newsvendor's risk preferences and they find subjects' order decisions systematically deviate from profit maximization. Therefore, studying newsvendor models with alternative risk preferences is important.

Recently, we have seen two growing bodies of literature that attempts to use alternative risk preferences to describe newsvendor decision biases from the risk-neutral newsvendor. First, within the *expected utility theory* (hereafter, EUT) framework in the field of classical economics, some researchers have studied the *risk-averse* newsvendor problem (e.g., Agrawal and Seshadri 2000; Eeckhoudt et al. 1995; and Wang et al. 2009). Second, within the *prospect theory* (hereafter, PT) framework in the field of behavioral economics, some researchers have studied the *loss-averse* newsvendor problem (e.g., Wang and Webster 2007, 2009; Wang 2010). As far as we know, there is a lack of unified framework for the newsvendor models with alternative preferences within these two frameworks. In view of this, the main purpose of this chapter is threefold: (1) we summarize the key research findings in newsvendor models with risk-averse and loss-averse preferences within EUT and PT frameworks and explain associated newsvendor decision biases; (2) we provide fairly comprehensive surveys of related newsvendor models within EUT and PT frameworks; and (3) we identify some directions for future research on the newsvendor models with alternative risk preferences within and beyond EUT and PT frameworks.

This chapter is organized as follows. In Sect. 7.2, we provide brief overviews of EUT and PT frameworks within which we introduce our definitions of risk aversion and loss aversion for studying the newsvendor problem. In Sect. 7.3, we study a general newsvendor model with the risk-averse preference in EUT and provide a survey of related research. In Sect. 7.4, we consider a general newsvendor model with the loss-averse preference in PT and provide a review of related research. Finally, in Sect. 7.5, we offer our concluding remarks and suggest opportunities for future research on the newsvendor problem with alternative risk preferences.

7.2 Overviews of Expected Utility Theory and Prospect Theory

In Sect. 7.2.1, we provide an overview of the EUT framework and introduce the definition of risk aversion. In Sect. 7.2.2, we provide an overview of the PT framework and introduce the definition of loss aversion.

7.2.1 Risk Aversion in EUT

EUT is concerned with choices among risky prospects with uncertain outcomes. It may be traced back to Bernoulli (1954) in response to the famous St. Petersburg paradox. Later, the development of EUT with a set of appealing axioms on preference by von Neumann and Morgenstern (1944) provided the basis for most subsequent analysis of decision-making behavior under uncertainty. Today, EUT is one of the fundamental theories in classical economics and applied fields such as finance, marketing, and operations management. For example, the classical newsvendor model is based upon risk neutrality, one of the risk preferences in EUT. We refer interested readers to Schoemaker (1982) for a comprehensive review of EUT.

According to EUT, risk preferences could be classified into three main categories: (1) a decision maker is *risk averse* if he prefers the expected monetary payoff over a lottery; (2) a decision maker is *risk loving* if he prefers the lottery over its expected monetary payoff; and (3) a decision maker is *risk neutral* if he is indifferent between a lottery and its expected monetary payoff. Since the risk-neutral newsvendor model is well known in the literature and the risk-taking newsvendor model is very rare, we are more interested in the risk aversion preference in EUT. Below we introduce a formal definition of risk aversion in EUT.

Definition 1 (Risk Aversion). Let $U(W)$ be the decision maker's utility function over the final wealth level W where $U(W)$ is twice differentiable. Then the decision maker is risk averse if the following two assumptions hold:

A1: $U'(W) > 0$ for all W.
A2: $U''(W) < 0$ for all W.

Assumption A1 implies that $U(W)$ is a strictly increasing function of W, meaning that more wealth is desirable. Assumption A2 implies the *diminishing marginal utility of wealth*, i.e., a dollar that helps us avoid poverty is more valuable than a dollar that helps us become very rich. Assumptions A1 and A2 are commonly used to describe the risk aversion preference in EUT (see, e.g., Arrow 1971; Pratt 1964).[1]

[1]Within the EUT framework, a risk-neutral decision maker's utility function $U(W)$ satisfies $U'(W) > 0$ and $U''(W) = 0$ for all W (i.e., a linear utility function), whereas a risk-loving decision

We let $r(W) = -U''(W)/U'(W)$ denote the well-known Arrow-Pratt measure of *absolute risk aversion* (Arrow 1971; Pratt 1964), which measures the insistence of an individual for more-than-fair bets. Risk-averse utility functions in EUT are commonly classified into three categories of absolute risk aversion: (1) *decreasing absolute risk aversion* (DARA), which states that as an individual becomes wealthier, he will be less risk averse (e.g., logarithmic, mixed exponential, and power utility functions), (2) *increasing absolute risk aversion* (IARA), which states that as an individual becomes wealthier, he will be more risk averse (e.g., quadratic utility function), and (3) *constant absolute risk aversion* (CARA), which states that an individual's degree of risk aversion is independent of the wealth level (e.g., exponential utility function). DARA, IARA, and CARA utility functions have been used to study the risk-averse newsvendor problem in the literature.

7.2.2 Loss Aversion in PT

In contrast with EUT that is commonly used in classical economics and applied fields to describe *rational* decision-making behavior under uncertainty, recently we have seen a fast development of behavioral economics that incorporates cognitive and psychological factors to describe *irrational* decision-making behavior under uncertainty. Probably the most well-known theory in behavioral economics is Kahneman and Tversky's (1979) prospect theory. PT states that people are (1) more sensitive to changes to a reference point (e.g., wealth) rather than absolute changes (i.e., reference dependence); (2) more averse to losses than attracted to same-sized gains (i.e., loss aversion); and (3) risk averse in the domain of gains and risk loving in the domain of losses (i.e., diminishing marginal sensitivity to changes with respect to the reference point). We refer interested readers to Bowman et al. (1999), Kahneman and Tversky (1979), and Köbberling and Wakker (2005) for comprehensive discussions of PT.

Unfortunately, the general form of PT often makes the newsvendor model intractable. In view of this, we focus on a simplified piecewise-linear form of PT defined below, and for simplicity, we refer to such a piecewise-linear PT utility function as *loss aversion*.

Definition 2 (Loss Aversion). Let $U(W)$ be the decision maker's utility function over the final wealth level W and let W_0 be the reference wealth level. The decision maker is loss averse if $U(W)$ is in the following piecewise-linear form:

$$U(W) = \begin{cases} W - W_0 & W \geq W_0 \\ \lambda(W - W_0), & W < W_0, \end{cases} \tag{7.1}$$

maker's utility function $U(W)$ satisfies $U'(W) > 0$ and $U''(W) > 0$ for all W (i.e., a convex utility function).

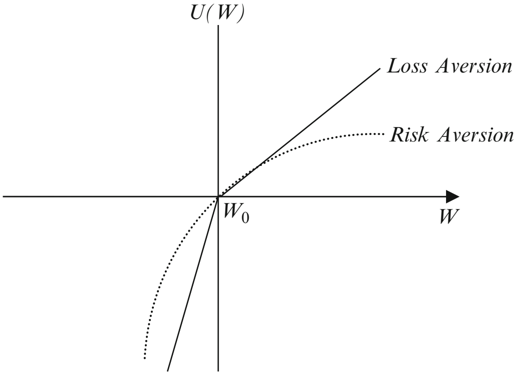


Fig. 7.1 Risk aversion versus loss aversion

where $\lambda > 1$ is defined as the loss aversion coefficient, and a higher value of λ implies a higher degree of loss aversion.

We note that the piecewise-linear loss-averse utility function defined above captures two key properties of PT (i.e., reference dependence and loss aversion), but it does not preserve the diminishing marginal sensitivity property of PT. Due to its simplicity, it has been commonly used by researchers in economics, finance, marketing, and operations management (e.g., Barberis and Huang 2001; Bell and Lattin 2000; Genesove and Mayer 2001; Wang and Webster 2009).

Figure 7.1 below shows the shapes of risk-averse (Definition 1) and loss-averse (Definition 2) utility functions to be used for our analyses of the newsvendor models in Sects. 7.3 and 7.4. In summary, loss aversion stems from behavioral economics and can be traced to the work of Kahneman and Tversky's (1979) prospect theory. It is distinguished from risk aversion in EUT by the presence of a reference point W_0 that determines whether an outcome is perceived as a loss or a gain, and by an abrupt change (kink) in the slope of the utility function at the reference point W_0.

7.3 Risk-Averse Newsvendor Models in EUT

In Sect. 7.3.1, we introduce our notation, assumptions, and other preliminaries based upon the classical risk-neutral newsvendor model. In Sect. 7.3.2, we consider a newsvendor model with an alternative risk-averse preference in EUT and summarize its key research findings. Finally, in Sect. 7.3.3, we provide a survey of related risk-averse newsvendor models.

7.3.1 The Classical Newsvendor Problem

In the classical newsvendor model setting, the newsvendor orders Q units of a short life-cycle product at a unit cost c from a supplier before the selling season begins and sells the product at a unit retail price $p > c$ during the selling season. Demand X is stochastic with PDF $f(x)$ and CDF $F(x)$ defined over the continuous interval I = [0, ∞).[2] As with most of the newsvendor models, we assume $F(x)$ is continuous, differentiable, invertible, and strictly increasing over I. We let $\bar{F}(x) = 1 - F(x)$ be the tail distribution. If realized demand x is higher than Q, then a unit shortage cost s is incurred on $x - Q$ units. While it is typical to assume $s \geq 0$ in the newsvendor literature, to be more general, we also allow the shortage cost s to be negative ($c - p < s < 0$). In other words, the case of $s > 0$ may account for a goodwill cost due to a stockout, whereas the case of $s < 0$ corresponds to situations where $x - Q$ units can be purchased and sold after demand is realized at a lower unit margin of $-s$ instead of $p - c$.[3] If realized demand x is lower than Q, then the newsvendor salvages $Q - x$ units of unsold products at a unit value $v < c$. The newsvendor has the following payoff function:

$$\pi(x,Q) = \begin{cases} \pi_-(x,Q) = px + v(Q-x) - cQ & x \leq Q \\ \pi_+(x,Q) = pQ - cQ - s(x-Q) & x > Q. \end{cases} \tag{7.2}$$

If the newsvendor is risk neutral, then there exists a unique optimal order quantity Q^N that satisfies the following first-order condition:

$$(p+s-c)\bar{F}(Q^N) - (c-v)F(Q^N) = 0. \tag{7.3}$$

For a risk-neutral newsvendor, the first term in (7.3) is the marginal benefit in *expected profit* if one more unit is ordered whereas the second term is the marginal cost in *expected profit* if one more unit is ordered. After conducting the comparative-statics analysis, the risk-neutral newsvendor's optimal order quantity is increasing in the selling price p, shortage cost penalty s, and salvage value v, but decreasing in the purchasing cost c. Since we are more interested in the newsvendor models with alternative risk preferences rather than risk neutrality, we refer interested readers to Khouja (1999), Lee and Nahmias (1990), Petruzzi and Dada (1999), Porteus (1990), and Qin et al. (2011) for comprehensive reviews of the risk-neutral newsvendor model and its extensions.

[2]To simplify analysis, we assume $I = [0, \infty)$. We refer interested readers to Wang et al. (2009) and Wang and Webster (2009) for newsvendor analyses based upon a more general demand assumption of $I = [a, b]$ where $a = 0$ and $b > a$.

[3]For example, Eeckhoudt et al. (1995) consider a newsvendor who is allowed to obtain additional newspapers at a cost c' satisfying $c < c' < p$. Thus, the newsvendor is still able to make money from replenishment orders to satisfy unmet demand (i.e., the shortage cost penalty is $s = c' - p \in (c-p, 0)$).

7.3.2 The Risk-Averse Newsvendor Problem

Consider a newsvendor with a risk-averse utility function $U(W)$ defined by Definition 1. After mapping the newsvendor's payoff function (7.2) into the risk aversion utility function $U(W)$, we can express the risk-averse newsvendor's expected utility function as follows:

$$E[U(\pi(X,Q))] = \int_0^Q U(\pi_-(x,Q))f(x)\mathrm{d}x + \int_Q^\infty U(\pi_+(x,Q))f(x)\mathrm{d}x. \tag{7.4}$$

The newsvendor will select an optimal order quantity Q^R to maximize the expected utility function $E[U(\pi(X,Q))]$. After taking the first and second derivatives of expression (7.4) with respect to Q, we get:

$$\begin{aligned}\mathrm{d}E[U(\pi(X,Q))]/\mathrm{d}Q = {} & (p+s-c)\int_Q^\infty U'(\pi_+(x,Q))f(x)\mathrm{d}x \\ & -(c-v)\int_0^Q U'(\pi_-(x,Q))f(x)\mathrm{d}x\end{aligned} \tag{7.5}$$

and

$$\begin{aligned}&\mathrm{d}^2E[U(\pi(X,Q))]/\mathrm{d}Q^2 \\ &\quad = (p+s-c)^2\int_Q^\infty U''(\pi_+(x,Q))f(x)\mathrm{d}x + (c-v)^2\int_0^Q U''(\pi_-(x,Q))f(x)\mathrm{d}x \\ &\quad\quad -(p+s-c)U'(\pi_+(Q,Q))\bar{F}(Q) - (c-v)U'(\pi_-(Q,Q))F(Q) < 0.\end{aligned} \tag{7.6}$$

From (7.5) and (7.6) we see that $\mathrm{d}E[U(\pi(X,0))]/\mathrm{d}Q > 0$, $\mathrm{d}E[U(\pi(X,\infty))]/\mathrm{d}Q < 0$, and $\mathrm{d}^2E[U(\pi(X,Q))]/\mathrm{d}Q^2 < 0$ for all $Q \in (0,\infty)$. Therefore, the risk-averse newsvendor's optimal order quantity $Q^R \in (0,\infty)$ is unique and satisfies the following first-order condition:

$$(p+s-c)\int_{Q^R}^\infty U'(\pi_+(x,Q^R))f(x)\mathrm{d}x - (c-v)\int_0^{Q^R} U'(\pi_-(x,Q^R))f(x)\mathrm{d}x = 0. \tag{7.7}$$

From (7.3) and (7.7), we can see some differences between the risk-neutral and risk-averse newsvendors' optimal order quantities. For the risk-averse newsvendor, the first term in (7.7) is the marginal benefit in *expected utility* due to an increase in the initial order quantity whereas the second term is the marginal loss in expected utility due to an increase in the initial order quantity. In other words, given that there is an underage with a probability of $\bar{F}(Q^R)$, if one more unit is ordered, then a unit underage cost c_u will be saved. However, in contrast with the risk-neutral newsvendor, there is also a decrease in the *marginal utility of underage* $U'(\pi_+)$ through a *wealth effect*, i.e., the diminishing marginal utility if the newsvendor is richer (by $U''(W) < 0$). Similarly, given that there is an overage with a probability of $F(Q^R)$, if one more unit is ordered, then a unit overage cost c_u will be incurred, and the decrease in wealth will be accompanied by an increase in the *marginal utility of overage* $U'(\pi_-)$ through a wealth effect.

We next use a numerical example to illustrate the results in the risk-averse newsvendor model analyzed above. We assume demand is uniformly distributed between [0, 100] and fix the newsvendor's purchasing cost at $c = \$100$. For simplicity of analysis, we let $s = v = W_0 = 0$ (i.e., the newsvendor's salvage value, shortage cost, and initial wealth are zero). Finally we assume the newsvendor has a CARA (exponential) utility function $U(W) = 1 - \mathrm{e}^{-rW}$, where $r > 0$ is the constant coefficient of absolute risk aversion.

After mapping the demand and utility functions into the risk-averse newsvendor's expected utility function (7.4) and plugging in the parameters, we can rewrite (7.4) as follows:

$$E[U(\pi(X,Q))] = \frac{(100-Q)(1-\mathrm{e}^{-r(p-100)Q})}{100} + \frac{rpQ + (\mathrm{e}^{-rpQ}-1)\mathrm{e}^{100rQ}}{100rp}.$$

Similarly, after applying some algebraic manipulations to the first-order condition (7.7), the risk-averse newsvendor's optimal order quantity Q^R satisfies the following first-order condition:

$$\left(\frac{p-100}{100}(100-Q^R) + \frac{1}{rp}\right)\mathrm{e}^{-rpQ^R} - \frac{1}{rp} = 0.$$

Figure 7.2 below gives us an example which shows the newsvendor's expected utility with respect to the order quantity Q when selling price $p = \$200$. It shows that the newsvendor's expected utility drops abruptly in the order quantity Q when the newsvendor is becoming more risk-averse (e.g., $r = 0.01$). Similarly, Table 7.1 below reports the risk averse newsvendor's optimal order quantities under various values of the selling price p and the coefficient of absolute risk aversion r. Note that the case of $r = 0$ corresponds to the classical newsvendor model based upon risk neutrality.

7.3.3 *Related Risk-Averse Newsvendor Models*

There is a large body of research on the risk-averse newsvendor problem in EUT. In this subsection, we mainly focus on newsvendor models with risk-averse utility functions in EUT and defined by Definition 1. We do not strictly organize those

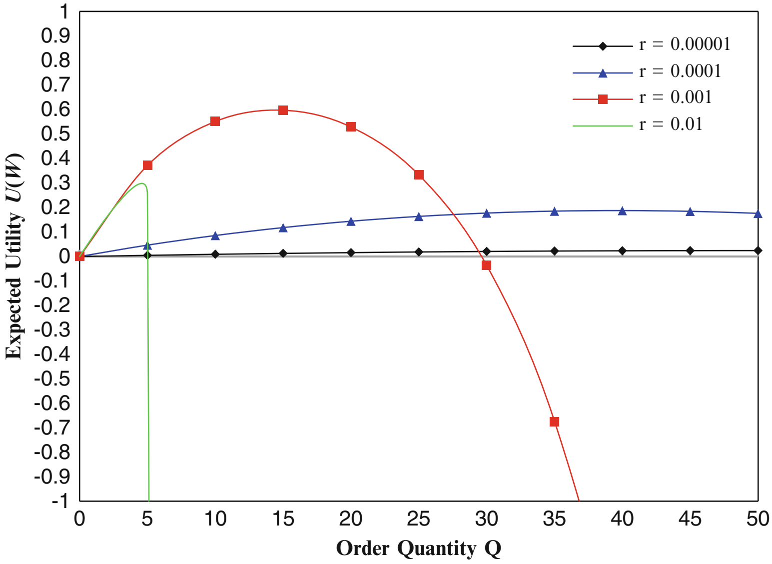


Fig. 7.2 Risk-averse newsvendor's expected utility with respect to the order quantity under various risk aversion coefficients r

Table 7.1 Risk-averse newsvendor's optimal order quantities under various selling price p and coefficients of absolute risk aversion r

P	$r = 0$	$r = 0.00001$	$r = 0.0001$	$r = 0.001$	$r = 0.01$
$200	50	49	40	14	3
$400	75	72	49	12	2
$600	83	80	47	9	1
$800	88	83	44	8	1
$1,000	90	85	40	7	1
$2,000	95	87	28	4	1
$4,000	98	83	18	2	0
$10,000	99	60	9	1	0
$40,000	100	24	3	0	0
$100,000	100	11	1	0	0
$400,000	100	4	0	0	0
$4,000,000	100	0	0	0	0

papers in the order of their publication time. Rather, we prefer to discuss more closely related papers together. We also note that we may omit some papers due to the limit of our knowledge of this part of literature.

As far as we know, Baron (1973) is the first to investigate a "newsvendor" type of decision-making problem under risk aversion. He shows that an increase in risk aversion (i.e., a higher value of the Arrow-Pratt measure of absolute risk aversion, $r(W)$) will decrease the optimal order quantity. However, as pointed out in

Eeckhoudt et al. (1995), Baron (1973) does not consider the newsvendor problem, per se, even though his analysis of a piecewise-linear payoff function bears the newsvendor form.

Eeckhoudt et al. (1995) extends Baron (1973) to a newsvendor model setting in which the newsvendor is allowed to obtain additional newspapers if the initial order quantity is insufficient to cover demand, and the newsvendor still makes money from replenishment orders. This assumption corresponds to the case of a negative shortage cost ($s < 0$) in our general newsvendor model setting in Sect. 7.3.2. Similar to Baron (1973), they find that a risk-averse newsvendor will order strictly less than a risk-neutral newsvendor, and the optimal order quantity decreases as the newsvendor becomes more risk averse. They also investigate the comparative-static effects of changes in price and cost parameters. They find that in contrast with the risk-neutral newsvendor model, the effects of selling price p and initial purchasing cost c on the risk-averse newsvendor's optimal order quantity are ambiguous, meaning that a risk-averse newsvendor's optimal order quantity may not increase in p and decrease in c as predicted by the risk-neutral newsvendor model. In addition, they investigate the comparative-static effects of changes in the independent *additive* background risk and mean-preserving demand risk on the risk-averse newsvendor's optimal order quantity.

Wang et al. (2009) extends Eeckhoudt et al. (1995) to a more general newsvendor model setting that allows for both positive and negative shortage cost. They derive the risk-averse newsvendor's optimal order quantity, but their main purpose is to characterize the relationship between a risk-averse newsvendor's optimal order quantity and selling price. For most commonly used CARA, IARA, and bounded DARA utility functions in EUT, they prove that a risk-averse newsvendor's optimal order quantity will decrease in selling price p if p is above a threshold price. For example, consider the numerical results in Sect. 7.3.2. From Table 7.1, we see that if the newsvendor is a little risk averse (e.g., $r = 0.0001$) and selling price p is relatively low (e.g., $p < \$400$), then the optimal order quantity is increasing in selling price. However, as the selling price becomes higher (e.g., $p > \$600$), the optimal order quantity begins to decrease in selling price and approaches zero when selling price is very high (e.g., $p = \$4,000,000$). In other words, the quantity that maximizes the newsvendor's expected utility at $r = 0.0001$ is 40 when the opportunity cost of a lost sale $(p - c)$ is only \$100, but when the opportunity cost is very high at \$3,999,900, the newsvendor's optimal order quantity does not increase, and in fact, goes to zero. As pointed out in Wang et al. (2009), such a counterintuitive result may be attributed to a limitation of EUT (Arrow 1971; Rabin 2000), i.e., *risk aversion within the EUT framework implies that people are approximately risk neutral when economic stakes are small.*

Next we briefly review other extensions of the risk-averse newsvendor problem. Agrawal and Seshadri (2000) investigate a risk-averse newsvendor model in which the selling price and order quantity decisions are made jointly. They find that a risk-averse newsvendor will charge a higher price and order less than the risk-neutral newsvendor if the demand distribution has the multiplicative form of relationship with price. Also, the risk-averse newsvendor will charge a lower price if the demand distribution has the additive form of relationship with price, but the effect on the quantity ordered depends on the demand sensitivity to selling price. Gaur and Seshadri (2005) consider a risk-averse newsvendor who hedges inventory risk when demand is correlated with the price of a financial asset. They show that a risk-averse newsvendor will order more inventory if he or she hedges the inventory risk. Keren and Pliskin (2006) consider a special risk-averse newsvendor model under uniform demand and derive the optimal order quantity in a closed form. Chen et al. (2007) extend Bouakiz and Sobel (1992) by studying a multi-period inventory model under risk aversion as well as multi-period models that coordinate inventory and pricing decisions. Van Mieghem (2007) studies resource diversification, flexibility, and/or demand pooling decisions in newsvendor networks featured with many products and many resources. He shows that a newsvendor with a general risk-averse utility function in EUT (and also under the mean-variance criteria) may invest more resources in certain networks than does a risk-neutral newsvendor. Sévi (2010) investigates a risk-averse newsvendor model with an independent *multiplicative* background risk and identifies conditions under which an introduction of the multiplicative background risk will decrease the risk-averse newsvendor's optimal order quantity. Choi and Ruszczynski (2011) consider a multi-product risk-averse newsvendor model with a CARA utility function. They establish a few basic properties for the newsvendor solution. One interesting result they find is that an increase in risk aversion does not always lead to a lower order quantity when demands of multiple products are strongly negatively correlated. Colombo and Labrecciosa (2012) consider the pricing decision for a risk-averse seller with a fixed supply. Demand uncertainty stems from random consumer valuations of the product. They show that if the distribution of consumer valuations exhibits an increasing generalized failure rate (IGFR) property, then the risk-averse seller will charge a lower price than a risk-neutral seller.

We wish to note that we focus on newsvendor models with risk-averse utility functions in EUT. An alternative approach to modeling risk aversion is mean-variance analysis (MV) due to Markowitz (1959), which has become a standard analytical tool for portfolio optimization. MV specifies a decision maker's objective as a function of the mean and variance of a performance measure such as profit. The MV framework is less precise and general than the EUT framework because the measure of "utility" depends on only two moments of the probability distribution of profit rather than the entire distribution, but as an important consequence, the framework generally affords greater analytical tractability and can be more readily implemented (e.g., mean and variance can be estimated even if the distribution is unknown). In addition, under certain conditions (e.g., normally distributed returns and CARA utility) MV and EUT are equivalent (i.e., return the same optimal decisions). MV has recently drawn much attention from researchers in operations management, especially with respect to the newsvendor problem, we refer interested readers to Anvari (1987), Chen and Federgruen (2000), Choi et al. (2008a, 2008b), Chung (1990), Gan et al. (2004), Lau (1980), Martínez-de-Albéniz and Simchi-Levi (2006), Van Mieghem (2007), Wei and Choi (2010), and references therein for this part of literature.

Finally, there are other risk-averse newsvendor models within other frameworks such as coherent measures of risk (e.g., Chen et al. 2009; Choi and Ruszczynski 2008) and value at risk (VaR, e.g., Gan et al. 2005). We refer interested readers to Choi et al. (2011) for a recent survey of related papers.

7.4 Loss-Averse Newsvendor Models in PT

In Sect. 7.4.1, we consider a newsvendor model with an alternative loss-averse preference in PT and summarize its key research findings. In Sect. 7.4.2, we provide a survey of related loss-averse newsvendor models.

7.4.1 The Loss-Averse Newsvendor Problem

Consider a newsvendor with a loss aversion utility function $U(W)$ in PT and defined by Definition 2. For simplicity of analysis, we assume the newsvendor's reference wealth level $W_0 = 0$. Similar analysis could be extended to the case of $W_0 \neq 0$. Following Lemma 1 in Wang and Webster (2009), we define $q_1(Q) = (\frac{c-v}{p-v})Q$ and $q_2(Q) = \left(\frac{p-c+s}{s}\right)Q$ as the loss-averse newsvendor's two breakeven quantities of realized demand, where if $x < q_1(Q)$ or $x > q_2(Q)$, then the newsvendor faces a loss (i.e., realized profit is negative) and if $q_1(Q) < x < q_2(Q)$, then the newsvendor faces a gain (i.e., realized profit is positive). After mapping the newsvendor's payoff function (7.2) into the loss aversion utility function (7.1), we can express the newsvendor's expected utility $E[U(\pi(X,Q))]$ as follows:

$$E[U(\pi(X,Q))] = E[\pi(X,Q)] + (\lambda - 1)\left(\int_0^{q_1(Q)} \pi_-(x,Q)f(x)\mathrm{d}x + \int_{q_2(Q)}^{\infty} \pi_+(x,Q)f(x)\mathrm{d}x\right). \quad (7.8)$$

In expression (7.8), the term $\int_0^{q_1(Q)} \pi_-(x,Q)f(x)\mathrm{d}x$ could be defined the expected *overage loss* meaning that the newsvendor earns a negative profit if realized demand is lower than $q_1(Q)$ and the term $\int_{q_2(Q)}^{\infty} \pi_+(x,Q)f(x)\mathrm{d}x$ could be defined as the expected *underage loss* meaning that the newsvendor earns a negative profit if realized demand is higher than $q_2(Q)$. The loss-averse newsvendor's expected utility is the expected profit plus the total expected underage and overage losses, biased by a factor of $\lambda - 1$. If $\lambda = 1$, then the newsvendor is risk neutral and the second term in (7.8) drops out.

After taking the first and second derivatives of $E[U(\pi(X,Q)]$ in (7.8) with respect to Q and applying some algebraic manipulations, we get:

$$\begin{aligned} \mathrm{d}E[U(\pi(X,Q)]/\mathrm{d}Q &= (p+s-c)[\bar{F}(Q)+(\lambda-1)\bar{F}(q_2(Q))] \\ &\quad -(c-v)[F(Q)+(\lambda-1)F(q_1(Q))] \end{aligned} \tag{7.9}$$

and

$$\begin{aligned} &\mathrm{d}^2E[U(\pi(X,Q)]/\mathrm{d}Q^2 \\ &\quad -(p+s-w)f(Q)-(\lambda-1)\left[\frac{(w-v)^2f(q_1(Q))}{p-v}+\frac{(p+s-w)^2f(q_2(Q))}{s}\right]. \end{aligned} \tag{7.10}$$

From (7.9) and (7.10), we see that $\mathrm{d}E[U(\pi(X,0))]/\mathrm{d}Q > 0$, $\mathrm{d}E[U(\pi(X,\infty))]/\mathrm{d}Q < 0$, and $\mathrm{d}^2E[U(\pi(X,Q))]/\mathrm{d}Q^2 < 0$ for all $Q \in (0,\infty)$. Then the loss-averse newsvendor's optimal order quantity $Q^L \in (0,\infty)$ is unique and satisfies the following first-order condition:

$$\begin{aligned} &(p+s-c)\bar{F}(Q^L)-(c-v)F(Q^L)+(\lambda-1)[(p+s-c)\bar{F}(q_2(Q^L)) \\ &\quad -(c-v)F(q_1(Q^L)]=0. \end{aligned} \tag{7.11}$$

From (7.11), we see that the first two terms reflects the classical risk-neutral newsvendor trade-off between the overage and underage costs. However, it is the third term $(\lambda-1)[(p+s-c)\bar{F}(q_2(Q^L))-(c-v)F(q_1(Q^L)]$ in (7.11) that causes the difference between the risk neutral and loss averse newsvendors' optimal order quantities. Define $(p+s-c)F(q_2(Q^L))$ as the loss-averse newsvendor's *marginal underage loss* and $(c-v)F(q_1(Q^L))$ as the loss-averse newsvendor's *marginal overage loss*. Then the third term in (7.11) reflects another type of tradeoff due to loss aversion. More specifically, it shows that if the loss-averse newsvendor's marginal underage loss is lower than the marginal overage loss, then loss aversion will lead to a larger order quantity than the risk-neutral newsvendor, whereas if the loss-averse newsvendor's marginal underage loss is higher than the marginal overage loss, then loss aversion will lead to a smaller order quantity than the risk-neutral newsvendor.

We next use exponential and normal demand distributions to illustrate our results. Figure 7.3 illustrates the loss-averse newsvendor's optimal order quantity for the choice of fixed parameters $\{p=1,\ c=0.5,\ v=0\}$, and an exponential demand distribution $f(x)=0.01\mathrm{e}^{-0.01x}$. As it shows, if shortage cost is low ($s \le 4$ to be exact), then the loss-averse newsvendor will order less than the risk-neutral newsvendor. In addition, the more loss-averse, the less the newsvendor's optimal order quantity. We also observe that if $s=0$, then the loss-averse newsvendor only orders 71%, 56%, and 47% of the risk-neutral newsvendor's optimal order quantity

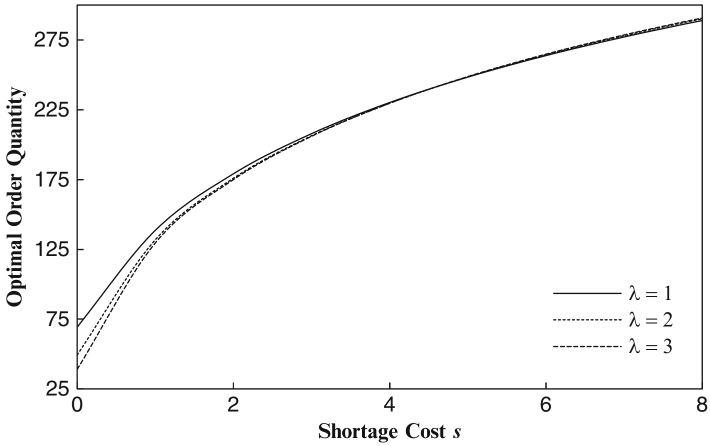


Fig. 7.3 Loss-averse newsvendor's optimal order quantity under exponential distribution

for $\lambda = 2$, $\lambda = 3$, and $\lambda = 4$, respectively. However, if shortage cost is high ($s \geq 5$ to be exact), then the loss averse newsvendor will order more than the risk-neutral newsvendor. In addition, the more loss averse, the more the newsvendor's optimal order quantity. We also observe that if $s = 8$, then the loss-averse newsvendor orders 100.4%, 100.5%, and 100.7% of the risk-neutral newsvendor's optimal order quantity for $\lambda = 2$, $\lambda = 3$, and $\lambda = 4$ respectively. In summary, if shortage cost is low, then the loss-averse newsvendor orders significantly less than the risk-neutral newsvendor; if shortage cost is high, then the loss-averse newsvendor orders slightly higher than the risk-neutral newsvendor.

Similarly, Fig. 7.4 illustrates the loss-averse newsvendor's optimal order quantity for the choice of fixed parameters $\{p = 1, c = 0.5, v = 0\}$, and a normal demand distribution with mean 100 and standard deviation 25. As it shows, if shortage cost is low ($s \leq 1$ to be exact), then the loss-averse newsvendor will order less than the risk-neutral newsvendor. In addition, the more loss averse, the less the newsvendor's optimal order quantity. We also observe that if $s = 0$, then the loss-averse newsvendor orders 99.31%, 98.16%, and 98.05% of the risk-neutral newsvendor's optimal order quantity for $\lambda = 2$, $\lambda = 3$, and $\lambda = 4$, respectively. However, if shortage cost is high ($s \geq 2$ to be exact), then the loss-averse newsvendor will order more than the risk-neutral newsvendor. In addition, the more loss-averse, the more the newsvendor's optimal order quantity. We also observe that if $s = 8$, then the loss-averse newsvendor orders 102.01%, 103.10%, and 103.80% of the risk-neutral newsvendor's optimal order quantity for $\lambda = 2$, $\lambda = 3$, and $\lambda = 4$ respectively. In summary, if shortage cost is low, then the loss-averse newsvendor orders slightly less than the risk-neutral newsvendor; if shortage cost is high, then the loss-averse newsvendor orders slightly higher than the risk-neutral newsvendor.

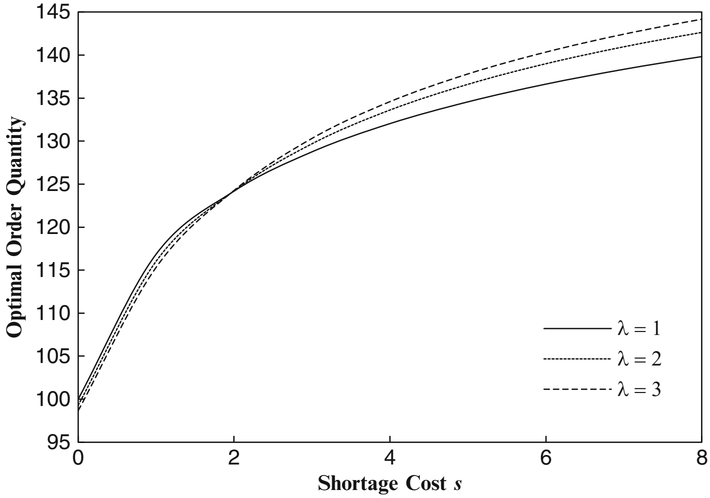


Fig. 7.4 Loss-averse newsvendor's optimal order quantity under normal distribution

Based upon these two numerical examples, we could classify short life-cycle products into two broad categories: the *low-shortage-cost* product and the *high-shortage-cost* product. We could list computers, food, and shirts as examples of low-shortage-cost products and airline meal, hotel rooms, and travel package as *high-shortage-cost* products. For example, in the airline industry, the cost per meal varied between \$2 and \$12 but the per meal shortage cost could be around \$120 (Goto et al. 2004). For high-shortage-cost products, a loss-averse newsvendor will order more than the risk-neutral newsvendor, whereas for low-shortage-cost products, a loss-averse newsvendor will order more than the risk-neutral newsvendor.

7.4.2 Related Loss-Averse Newsvendor Models

Compared to the large body of research on the risk-averse newsvendor problem within the EUT framework, there is much less research on the loss-averse newsvendor problem within the PT framework. As far as we know, Schweitzer and Cachon (2000) and Wang and Webster (2007, 2009) are among the first studying the loss-averse newsvendor problem.

Schweitzer and Cachon (2000) conduct two experiments to test the newsvendor problem with alternative risk preferences, including risk neutrality, risk aversion, and risk-loving preferences in EUT, and loss aversion in PT, among others. In their analysis of a loss-averse newsvendor problem without a shortage cost (i.e., $s = 0$), they find that a loss-averse newsvendor's optimal order quantity is less than a risk-neutral newsvendor and decreasing in the loss aversion level λ.

Wang and Webster (2009) study a more general loss-averse newsvendor problem with a positive shortage cost ($s \geq 0$). They establish necessary and sufficient conditions under which a loss-averse newsvendor will order (less) more than a risk-neutral newsvendor and the loss-averse newsvendor's optimal order quantity will increase (decrease) in the loss aversion level λ. They also show that a loss-averse newsvendor's optimal order quantity may increase in purchasing cost and decrease in selling price, which can never occur in the risk-neutral newsvendor model.

Wang and Webster (2007) study a decentralized supply chain setting in which a single risk-neutral manufacturer is selling a perishable product to a single loss-averse retailer facing uncertain demand.[4] They investigate the role of a gain/loss sharing provision for mitigating the loss aversion effect, and design a distribution-free gain/loss sharing-and-buyback contract that can coordinate the supply chain.

Wang (2010) extends the loss-averse newsvendor problem to a game setting where multiple loss-averse newsvendors are competing for inventory from a risk-neutral supplier. They show that under the proportional demand allocation rule by the supplier, there exists a unique Nash equilibrium order quantity in the loss-averse newsvendor game. They also find that if *loss aversion effect* is strong enough (i.e., λ is large enough), then the total order quantity of the loss-averse newsvendors in the decentralized supply chain will be less than that of an integrated supply chain. This result contrasts sharply with the result in the risk-neutral newsvendor game in which the total order quantity of the risk-neutral newsvendors in the decentralized supply chain will be more than that of an integrated supply chain due to *demand stealing effect* (see e.g., Lippman and McCardle 1997 and Cachon 2003).

Finally, Geng et al. (2010) consider a single-period newsvendor model with a general PT utility function and exponential demand and characterize the optimal inventory decision. Shen et al. (2011) study a loss-averse newsvendor-like manufacturer who purchases a component from a supplier under a wholesale price-only contract and an uncertain spot purchase price. They find that the purchasing decision of the loss-averse manufacturer differs from those of risk-neutral and risk-averse manufacturers.

7.5 Conclusions and Future Research

In this chapter, we investigate newsvendor models with alternative risk preferences within the EUT and PT frameworks. Based upon a general newsvendor model setting, we characterize the optimal order quantities of risk-averse and loss-averse newsvendors and discuss associated decision biases from the risk-neutral newsvendor model. We also provide surveys of related risk-averse and loss-averse

[4] Wang and Webster (2009) and Wang and Webster (2007) are based upon Chap. 3 and 4 of Wang's (2003) doctoral dissertation.

newsvendor models in EUT and PT. We believe that studying newsvendor models with alternative risk preferences are becoming important.

There are two worthwhile directions for future research related to the newsvendor problem with alternative risk preferences. First, most of the newsvendor research is based upon analytical models with assumed risk preferences from classical or behavioral economic theories. As a result, those theoretical newsvendor models only tell managers what actions they should take if their risk preferences are consistent with model assumptions, but little is known about the real risk preferences of the managers making newsvendor decisions. Therefore, there is a need to test the risk preferences of managers by experiments (e.g., Becker-Peth et al. 2011; Bolton and Katok 2008; and Schweitzer and Cachon 2000) or by surveys of managers about their risk preferences when they are making newsvendor decisions (e.g., Corbett and Fransoo 2007). Second, in addition to risk preferences within EUT and PT frameworks, there is a need to study the newsvendor problem with alternative risk preferences and objectives such as regret theory (e.g., Engelbrecht-Wiggans and Katok 2008), fairness (e.g., Cui et al. 2007; Loch and Wu 2008), bounded rationality (e.g., Su 2008), and achieving a profit and/or revenue target (He and Khouja 2011; Lau and Lau 1988; Yang et al. 2011). We refer interested readers to Bendoly et al. (2006), Gino and Pisano (2008), and Loch and Wu (2007) for detailed surveys of other behavioral frameworks.

References

Agrawal, V., & Seshadri, S. (2000). Impact of uncertainty and risk aversion on price and order quantity in the newsvendor problem. *Manufacturing & Service Operations Management,* 2, 410–423.

Anvari, M. (1987). Optimality criteria and risk in inventory models: the case of the newsboy problem. *The Journal of the Operational Research Society,* 38, 625–632.

Arrow, K. J. (1971). *Essays in the theory of risk-bearing.* Chicago, IL: Markham Publishing Company.

Barberis, N., & Huang, M. (2001). Mental accounting, loss aversion, and individual stock returns. *Journal of Finance,* 56, 1247–1292.

Baron, D. P. (1973). Point estimation and risk preferences. *Journal of the American Statistical Association,* 68, 944–950.

Becker-Peth, M., Katok, E., & Thonemann, U. W. (2011). Designing contracts for irrational but predictable newsvendors. Available at SSRN: http://ssrn.com/abstract=1567614. Accessed 31 July 2011.

Bell, D. R., & Lattin, J. M. (2000). Looking for loss aversion in scanner panel data: the confounding effect of price response heterogeneity. *Marketing Science,* 19, 185–200.

Bendoly, E., Donohue, K., & Schultz, K. L. (2006). Behavior in operations management: assessing recent findings and revisiting old assumptions. *Journal of Operations Management,* 24, 737–752.

Bernoulli, D. (1954). Exposition of a new theory on the measurement of risk (original 1738). *Econometrica,* 22, 23–26.

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

Bouakiz, M., & Sobel, M. J. (1992). Inventory control with an exponential utility criterion. *Operations Research,* 40, 603–608.

Bowman D., Minehart, D., & Rabin, M. (1999). Loss aversion in a consumption savings model. *Journal of Economic Behavior & Organization,* 38, 155–178.

Cachon, G. P. (2003). Supply chain coordination with contracts. In: A. G. de Kok & S. C. Graves (Eds.), *Handbook in operations research and management science, volume on supply chain management: design, coordination and operation* (PP. 229–339). Amsterdam: North Holland.

Chen, F., & Federgruen, A. (2000). Mean-variance analysis of basic inventory models. Working paper, Columbia University, New York.

Chen, X., Sim, M., Simchi-Levi, D., & Sun, P. (2007). Risk aversion in inventory management. *Operations Research,* 55, 828–842.

Chen, Y., Xu, M., & Zhang, Z. (2009). A risk-averse newsvendor model under the CVaR criterion. *Operations Research,* 57, 1040–1044.

Choi, S., & Ruszczynski, A. (2008). A risk-averse newsvendor with law invariant coherent measures of risk. *Operations Research Letters,* 36, 77–82.

Choi, S., & Ruszczynski, A. (2011). A multi-product risk-averse newsvendor with exponential utility function. *European Journal of Operational Research,* 214, 78–84.

Choi, S., Ruszczynski, A., & Zhao, Y. (2011). A multiproduct risk-averse newsvendor with law-invariant coherent measures of risk. *Operations Research,* 59, 346–364.

Choi, T. M., Li, D., & Yan, H. (2008a). Mean-variance analysis of a single supplier and retailer supply chain under a returns policy. *European Journal of Operational Research,* 184, 356–376.

Choi, T. M., Li, D., Yan, H., & Chiu, C. H. (2008b). Channel coordination in supply chains with agents having mean-variance objectives. *Omega,* 36, 565–576.

Chung, K. H. (1990). Risk in inventory models: the case of the newboy problem – optimality conditions. *The Journal of the Operational Research Society,* 41, 173–176.

Colombo, L. & Labrecciosa, P. (2012). A note on pricing with risk aversion. *European Journal of Operational Research,* 216, 252–254.

Corbett, C. J., & Fransoo, J. C. (2007). Entrepreneurs and newsvendors: do small businesses follow the newsvendor logic when making inventory decisions? Available at SSRN: http://ssrn.com/abstract=1009330. Accessed 31 July 2011.

Cui, Tn.H., Raju, J. S., & Zhang, Z. J. (2007). Fairness and channel coordination. *Management Science,* 53, 1303–1314.

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

Engelbrecht-Wiggans, R., & Katok, E. (2008). Regret and feedback information in first-price sealed-bid Auctions. *Management Science,* 54, 808–819.

Fisher, M., & Raman, A. (1996). Reducing the cost of demand uncertainty through accurate response to early sales. *Operations Research,* 44, 87–99.

Gan, X., Sethi, S. P., & Yan, H. (2004). Coordination of supply chain with risk-averse agents. *Production and Operations Management,* 13, 135–149.

Gan, X., Sethi, S. P., & Yan, H. (2005). Channel coordination with a risk-neutral supplier and a downside-risk-averse retailer. *Production and Operations Management,* 13, 135–149.

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

Geng, W., Zhao, X., & Gao, D. (2010). A single-period inventory system with a general S-shaped utility and exponential demand. *Journal of Systems Science and Systems Engineering,* 19, 227–236.

Genesove, D., & Mayer, C. (2001). Loss aversion and seller behavior: evidence from the housing market. *Quarterly Journal of Economics,* 116, 1233–1260.

Gino, F., & Pisano, G. (2008). Toward a theory of behavioral operations. *Manufacturing & Service Operations Management,* 10, 676–691.

Goto, J. H., Lewis, M. E., & Puterman, M. L. (2004). Coffee, tee, or . . .?: a Markov decision process model for airline meal provisioning. *Transportation Science,* 38, 107–118.

He, X., & Khouja, M. (2011). Pareto analysis of supply chain contracts under satisficing objectives. *European Journal of Operational Research,* 214, 53–66.

Kahneman, D., & Tversky, A. (1979). Prospect Theory: an analysis of decisions under risk. *Econometrica,* 47, 263–291.

Keren, B., & Pliskin, J. S. (2006). A benchmark solution for the risk-averse newsvendor problem. *European Journal of Operational Research,* 174, 1643–1650.

Khouja, M. (1999). The single-period (news-vendor) problem: literature review and suggestions for future research. *Omega,* 27, 537–553.

Köbberling, V., & Wakker, P. P. (2005). An index of loss aversion. *Journal of Economic Theory,* 122, 119–131.

Lau, H. S. (1980). The Newsboy problem under alternative optimization objectives. *The Journal of the Operational Research Society,* 31, 525–535.

Lau, H. S., & Lau, A. H. L. (1988). Maximizing the probability of achieving a target profit in a two-product newsboy problem. *Decision Sciences,* 19, 392–408.

Lippman, S. A., & McCardle, K. F. (1997). The competitive newsboy. *Operations Research,* 45, 54–65.

Loch, C. H., & Wu, Y. (2007). Behavioral operations management. *Foundations and Trends in Technology, Information and Operations Management,* 1(3), 121–232.

Loch, C. H., & Wu, Y. (2008). Social preferences and supply chain performance: an experimental study. *Management Science,* 54, 6–38.

Markowitz, H. (1959). Portfolio selection: efficient diversification of investment.Cowls foundation monograph 16, New Haven:Yale University Press.

Martínez-de-Albéniz, V., & Simchi-Levi, D. (2006). Mean-variance trade-offs in supply contracts. it Naval Research Logistics, 53, 603–616.

Patsuris, P. (2001). Christmas sales: the worst growth in 33 years. *Forbes.* http://www.forbes.com/2001/10/30/1030retail.html. Accessed 30 October 2001.

Petruzzi, N. C., & Dada, M. (1999). Pricing and the newsvendor problem: a review with extensions. *Operations Research,* 47, 183–194.

Porteus, E. L. (1990). Stochastic inventory model. In S. C. Graves, A. H. G. R. Kan, & P. H. Zipkin (Eds.), *Handbook in Operations Research and Management Science (pp. 605–652).* Amsterdam: North Holland.

Pratt, J. W. (1964). Risk aversion in the small and in the large. *Econometrica,* 32, 122–136.

Qin, Y., Wang, R., Vakharia, A. J., Chen, Y., Michelle, M. H., & Seref, M. M. H. (2011). The newsvendor problem: review and directions for future research. *European Journal of Operational Research,* 213, 361–374.

Rabin, M. (2000). Risk aversion and expected-utility theory: a calibration theorem. *Econometrica,* 68, 1281–1292.

Schoemaker, P. (1982). The expected utility models: its variants, purposes, evidence and limitations. *Journal of Economic Literature,* 20, 529–563.

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

Sévi, B. (2010). The newsvendor problem under multiplicative background risk. *European Journal of Operational Research,* 200, 918–923.

Shen, H., Pang, Z., & Cheng, T. C. E. (2011). The component procurement problem for the loss-averse manufacturer with spot purchase. *International Journal of Production Economics,* 132, 146–153.

Su, X. (2008). Bounded rationality in newsvendor models. *Manufacturing & Service Operations Management,* 10, 566–586.

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

von Neumann, J., & Morgenstern, O. (1944). *The theory of games and economic behavior.* Princeton, NJ: Princeton University Press.

Wang, C. X. (2003). Decision bias in the newsvendor problem ans supply chains. Doctoral disseration, Syracuse University Whitman School of Management, Syracuse, NY.

Wang, C. X. (2010). The loss-averse newsvendor game. *International Journal of Production Economics,* 124, 448–452.
Wang, C. X., & Webster, S. (2007). Channel coordination for a supply chain with a risk-neutral manufacturer and a loss-averse retailer. *Decision Sciences,* 38, 361–389.
Wang, C. X., & Webster, S. (2009). The loss-averse newsvendor problem. *Omega,* 37, 93–105.
Wang, C. X., Webster, S., & Suresh, N. C. (2009). Would a risk-averse newsvendor order less at a higher selling price? *European Journal of Operational Research,* 196, 544–553.
Wei, Y., & Choi, T. M., 2010. Mean-variance analysis of supply chains under wholesale pricing and profit sharing schemes. *European Journal of Operational Research,* 204, 255–262.
Yang, S., Shi, C. V., & Zhao, X. (2011). Optimal ordering and pricing decisions for a target oriented newsvendor. *Omega,* 39, 110–115.

Chapter 8
Newsvendor Problems with VaR and CVaR Consideration

Werner Jammernegg and Peter Kischka

Abstract In this chapter, we consider approaches to express the risk preferences of a newsvendor by means of the risk measures value at risk (VaR), conditional value at risk (CVaR), and the mean-CVaR rule, which usually is defined as a convex combination of expected profit and CVaR. With these risk measures the decision maker can exploit risk-averse or risk-neutral behavior. In addition, we introduce a more general mean-CVaR measure where also risk- taking behavior can be expressed. The overall goal of the paper is a comparative analysis of these risk measures in the newsvendor framework. On the one hand VaR, CVaR and the (general) mean-CVaR, measures are used as objective functions to derive the respective optimal order quantity. Extensions of the basic models are reviewed. On the other hand the risk measures, especially VaR, are constraints of the model. We first review models with the expected profit as objective. Then the general mean-CVaR measure is taken as objective function and a service constraint and a loss constraint are added. In this framework, the risk attitudes of the newsvendor can be deduced from the characteristics of a product together with the specified service target and loss target.

Keywords CVaR • VaR • Risk • Mean-CVaR measure • Convex combination • Service and loss constraints

---

W. Jammernegg (✉)
Department of Information Systems and Operations, WU Vienna University of Economics and Business, Nordbergstraße 15, A-1090 Vienna, Austria
e-mail: werner.jammernegg@wu.ac.at

P. Kischka
Department of Business Statistics, Friedrich Schiller University Jena, Carl-Zeiß-Straße 3, 07743 Jena, Germany
e-mail: p.kischka@wiwi.uni-jena.de

T.-M. Choi (ed.), *Handbook of Newsvendor Problems*, International Series in Operations Research & Management Science 176, DOI 10.1007/978-1-4614-3600-3_8,

8.1 Introduction

Historically, in inventory management the newsvendor model is formulated with the objective to maximize the expected profit. Later on important streams of research suggest, e.g., to maximize the expected utility of profit (see Eeckhoudt et al. 1995) or to use the mean-variance objective (Lau 1980; Choi and Chiu 2012). There is some critique from a theoretical and/or from an empirical point of view on these (and other) approaches and there are still new suggestions for objective functions in the literature (see, e.g., Gao et al. 2011). Another early objective function is the probability to exceed a specific target profit (Lau 1980); this approach is closely related to the value at risk (VaR) measure which is a central concept for this chapter.

VaR and CVaR (conditional value at risk) are risk measures originating in the theory of finance. They are used in the newsvendor context to express and to formulate the risk attitudes of the decision maker. The CVaR measures the expected profit falling below a quantile level of the profit distribution known as VaR. In literature, these risk measures are not only used as objective functions but also as constraints. CVaR is a coherent risk measure, which is an important decision-theoretical feature (see, e.g., Artzner et al. 1999); this property does neither hold for VaR nor for the mean-variance measure.

A more general class of risk measures are the so-called mean-deviation rules. An example is of course the mean-variance measure. Recently, the CVaR is also used as a deviation measure in inventory management. Usually, a convex combination of the expected profit and the CVaR for a specified quantile level is considered; then the mean-CVaR model is coherent. In this framework, risk-neutral as well as risk-averse attitudes of the decision maker can be expressed. In general, this means that the optimal order quantity of a risk-averse newsvendor is smaller than that of a risk-neutral, profit-maximizing decision maker. In the sense of Fisher (1997) this seems reasonable for functional products with focus on cost minimization. But for innovative products, the focus should be on high levels of product availability where a risk-taking behaviour—the random profit is preferred to the expected profit—is more appropriate. Thus, a general mean-CVaR measure is used where the decision maker can exploit risk-averse, risk-neutral, as well as risk-taking behavior. For a recent review of newsvendor models including risk preferences of the decision maker, we refer to Li et al. (2011), Sect. 1.1 and Qin et al. (2011).

The overall goal of this chapter is a comparative analysis of VaR, CVaR, and mean-CVaR as objectives and as constraints, respectively, in newsvendor models. We start the analysis with preliminaries stating the notation and the basic results of the classical, risk-neutral newsvendor model in Sect. 8.2. In Sects. 8.3–8.5, the risk measures are used to formulate the objective function. Beside the basic model we refer to some extensions, e.g., concerning multi-product and price-setting newsvendor models. The third section is devoted to the VaR objective and the related objective to maximize the probability of exceeding a specified target profit. Then the CVaR criterion is presented in Sect. 8.4 and the CVaR-optimal order quantity is derived. In the fifth section mean-deviation rules are discussed, especially the mean-CVaR objective. We present the optimal order quantity depending on the risk parameters and compare in a numerical example the profit functions and the optimal order quantities of risk averse, risk-neutral, and risk-taking decision makers.

In Sect. 8.6, some risk measures are used as constraints. E.g., to avoid low profits or even high losses. First, VaR is used as constraint when the objective is to maximize expected profit. Then a general model is formulated using the general mean-CVaR measure of Sect. 8.5 as objective function. In addition, two constraints are added. The loss constraint is a specific VaR—constraint which is specified by an upper bound for the probability to result in losses. Moreover, the service constraint defines a lower bound for the cycle service level, i.e., the probability not to run out of stock. Here, the optimal order quantity is given by a two-sided control limit policy depending on the risk parameters. The characteristics of the product together with the loss target and the service target provide information to specify the risk preferences of the inventory manager with respect to the product under consideration.

In Sect. 8.7, we discuss some recent generalisations concerning the mean-CVaR rule as objective function. Finally, the basic intentions of the chapter are summarized in Sect. 8.8.

8.2 Preliminaries

We introduce our notation for the classical single-product newsvendor model.

Let X denote the random demand with nonnegative support and let p, c, z be the per unit selling price, purchase cost, and salvage value, respectively. We assume $p > c > z$.

Let y be the order quantity. Then the random profit is given by

$$g(y,X) = (p-c)\,y - (p-z)\,(y-X)^+ \tag{8.1}$$

with $(y-X)^+ = \max(y-X,0)$.

Let F denote the distribution function of X; we assume that F^{-1} exists. It is well known that the solution of

$$\max_y E(g(y,X)) \tag{8.2}$$

is given by

$$y^* = F^{-1}\left(\frac{p-c}{p-z}\right). \tag{8.3}$$

Even if F is invertible the distribution function F_y of the profit (8.1) has a point of discontinuity at $(p-c)y$, more precisely (Jammernegg and Kischka 2007, see Fig. 8.1)

$$F_y(t) = \begin{cases} F\left(y + \frac{t-(p-c)y}{(p-z)}\right) & \text{for } t < (p-c)y \\ 1 & \phantom{\text{for }} t \geq (p-c)y \end{cases}. \tag{8.4}$$

We have

$$F(y) = \sup\left\{F_y(t)|t < (p-c)y\right\}. \tag{8.5}$$

In the following, we consider risk measures as objectives or constraints for the classical newsvendor model. Extensions of the basic model, e.g., by including shortage cost or price-dependent demand, are mentioned in the respective sections.

8.3 Value at Risk Criterion

8.3.1 General Definition

Let Z be some profit variable with distribution function F_Z and let $\alpha \in [0,1]$.

The value at risk of Z is

$$\text{VaR}_\alpha(Z) = \inf\{z|F_Z(z) \geq \alpha\}. \tag{8.6}$$

If F_Z is continuous at z_0 and strictly increasing in a neighborhood of z_0 we have for $\alpha = F(z_0)$

$$\text{VaR}_\alpha(Z) = F_Z^{-1}(\alpha). \tag{8.7}$$

The VaR is a widespread measure of risk in finance: The probability that the profit Z is below $\text{VaR}_\alpha(Z)$ equals the prescribed α. Note that in decision theory, -VaR often is denoted as a measure of risk whereas VaR is denoted as a preference functional.

8.3.2 Newsvendor Model With VaR Objective

Let $g(y,X)$ be the profit in the newsvendor model (see (8.1)).

From (8.4) and (8.5), immediately follows:

$$\begin{aligned}&\text{For } \alpha < F(y): \text{VaR}_\alpha(g(y,X)) = F_y^{-1}(\alpha) = F^{-1}(\alpha)(p-z)-(c-z)y.\\&\text{For } \alpha \geq F(y): \text{VaR}_\alpha(g(y,X)) = (p-c)y.\end{aligned} \tag{8.8}$$

Moreover, for $\alpha < F(y): P(g(y,X) \leq \text{VaR}_\alpha(g(y,X)) = \alpha$.

The newsvendor problem with the objective VaR is given by

$$\max_y \text{VaR}_\alpha(g(y,X)). \tag{8.9}$$

This objective can also be interpreted as the maximization of the probability to achieve a given target (see (8.11) and (8.12) and the following discussion).

As can be seen from (8.8), VaR_α is an increasing linear function in y for $y \leq F^{-1}(\alpha)$ and a decreasing linear function in y for $y > F^{-1}(\alpha)$. Thus, the VaR-optimal order quantity $y^*(\alpha)$ is given by

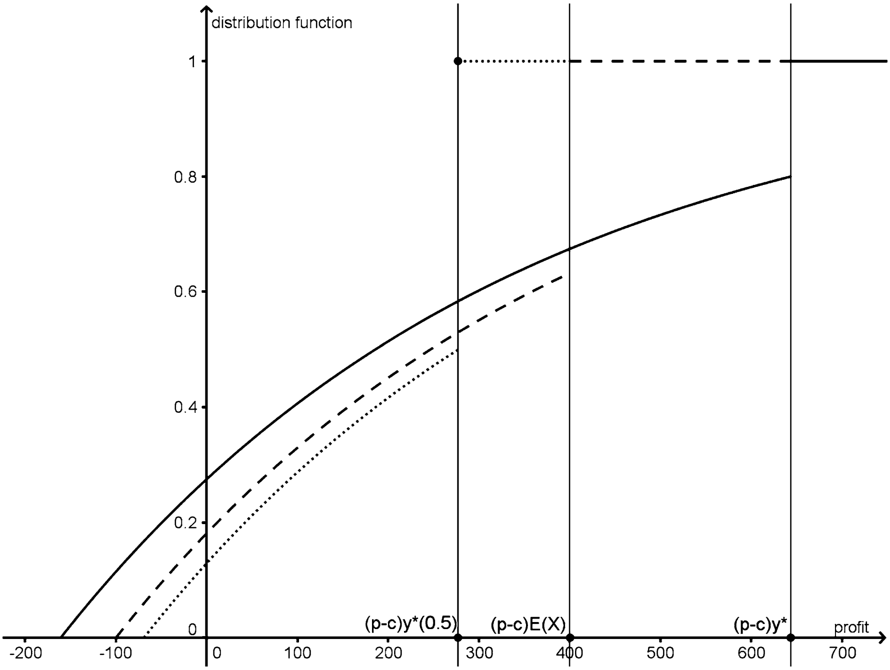


Fig. 8.1 Distribution functions of profit for VaR newsvendor (*dotted line*), expected demand newsvendor (*dashed line*), and classical newsvendor (*solid line*)

$$\arg\max_{y} VaR(g(y,X)) = y^*(\alpha) = F^{-1}(\alpha). \tag{8.10}$$

Note that the optimal solution is independent of p, c, z; it only depends on F and the prescribed α. In Chiu and Choi (2010), a price-setting newsvendor problem is considered with the value at risk as objective function. There the optimal order quantity depends on the stochastic part of demand and on the optimal price only via c. Another (single-period) inventory model with VaR objective is discussed in Tapiero (2005).

We illustrate the previous analysis by comparing the distribution functions of the profit for the classical newsvendor, the VaR newsvendor, and the newsvendor ordering the expected demand. The random demand is exponentially distributed with expected demand $E(X) = 100$ units; furthermore, $p = 10$, $c = 6$, $z = 5$, and $\alpha = 0.5$.

Depending on the respective optimal order quantity y for the different objectives in Fig. 8.1, the intervals of possible profits $[(z-c)y,\ (p-c)y]$ are shown.

Compared with the classical newsvendor, the probability of resulting in loss is just about half the amount for the VaR newsvendor. But on the other hand, the maximum VaR profit is only 277 currency units whereas that of the classical newsvendor is 643 currency units. This trade-off is a significant explanation why an inventory manager in praxis often pursues the pull-to-center strategy, i.e., they order the expected demand (the maximum expected demand profit is 400 currency units). In addition, small order quantities result in low levels of customer service which in any case is not advantageous for products with high profit value; remember that $y^*(\alpha)$ is independent of p, c, and z.

Up to now the value α is prescribed and the corresponding profit for an order quantity y is computed, which is to be maximized. Alternatively, one can prescribe a target profit B and look for an order quantity such that the probability of exceeding B is maximized. The seminal paper of this approach is Lau (1980). The formal problem and its solution is

$$\max_y P(g(y,X) \geq B) \tag{8.11}$$

$$\arg\max_y P(g(y,X) \geq B) = y^*(B) = \frac{B}{p-c}. \tag{8.12}$$

Note that the optimal solution is independent of F and z. In case of positive shortage cost, the optimal order quantity depends also on the demand distribution (Lau 1980, p. 531).

The corresponding maximal probability is

$$P(g(y^*(B),X) \geq B) = 1 - F(y^*(B))$$
$$P(g(y^*(B),X) < B) = F(y^*(B)).$$

Note that (see (8.4))

$$P(g(y^*(B),X) \leq B) = 1.$$

There are also models where the target profit B is not assumed to be fixed but depends on the order quantity y; an example is the expected profit $E(g(y,X))$ (see Parlar and Weng 2003). Shi and Chen (2007) show that for objective (8.11) the wholesale price contract is Pareto-optimal which does not hold for the expected profit criterion. For a price-setting newsvendor, in addition to a target profit, a target revenue is considered leading to a model with two objective functions (see Yang et al. 2011).

The approaches (8.9), (8.10) and (8.11), (8.12) are closely related.

Define for a given satisfying profit B

$$\alpha := F\left(\frac{B}{p-c}\right).$$

Then from (8.10), we have

$$y^*(\alpha) = F^{-1}(\alpha) = \frac{B}{p-c} = y^*(B)$$

and therefore

$$\mathrm{VaR}_\alpha(g(y^*(B),X)) = (p-c)F^{-1}(\alpha) = B.$$

Conversely, for a given α define

$$B := (p-c)F^{-1}(\alpha).$$

Then from (8.12), we have

$$y^*(B) = \frac{B}{p-c} = F^{-1}(\alpha) = y^*(\alpha)$$

and therefore

$$P(g(y^*(\alpha),X) \geq B) = 1 - F(y^*(B)) = 1 - \alpha.$$

Summarizing, from the results obtained so far it is evident that the optimal VaR order quantity is independent of the selling price, the purchase cost, and the salvage value whereas the optimal order quantity for the target profit newsvendor does not depend on the demand distribution. The risk measures used in the following objective functions do not result in optimal decisions that show these deficiencies.

8.4 Conditional Value at Risk criterion

8.4.1 General Definition

There are different possibilities to define the conditional value at risk (see, e.g., Acerbi and Tasche 2002; Rockafellar and Uryasev 2000).

For $\beta \in [0,1]$, the generalized (lower) inverse function of the distribution function F_Z of a random variable Z is defined by

$$F_Z^*(\beta) = \inf\{z | F_Z(z) \geq \beta\} \qquad (0 < \beta \leq 1)$$
$$F_Z^*(0) = \lim_{\beta \to 0} F_Z^*(\beta).$$

The conditional value at risk of Z given $\alpha \in [0,1]$ is

$$\mathrm{CVaR}_\alpha(Z) := \frac{1}{\alpha}\int_0^\alpha F_Z^*(\beta)\mathrm{d}\beta. \tag{8.13}$$

Alternatively, the CVaR_α can be defined by the generalized upper inverse function.

From Rockafellar and Uryasev (2000), we have

$$\mathrm{CVaR}_\alpha(Z) = \sup_{t\in\mathbb{R}}\left(t - \frac{1}{\alpha}E(t-Z)^+\right). \tag{8.14}$$

If F_Z^{-1} exists, we have

$$\begin{aligned}\mathrm{CVaR}_\alpha(Z) &= E(Z|Z \leq F_Z^{-1}(\alpha)) \\ &= E(Z|Z \leq \mathrm{VaR}_\alpha(Z)).\end{aligned} \tag{8.15}$$

Note again that in decision theory -CVaR often is denoted as a measure of risk whereas CVaR is denoted as a preference functional.

8.4.2 Newsvendor Model With CVaR Objective

For $Z = g(y,X)$, we have (see Jammernegg and Kischka 2007, p. 108)

$$F_Z^*(\beta) = F_y^*(\beta) = \begin{cases} F_y^{-1}(\beta) & \text{for } \beta < F(y) \\ (p-c)y & \beta \geq F(y) \end{cases}$$

and therefore:

$$\begin{aligned}\text{For } \alpha < F(y): \quad \mathrm{CVaR}_\alpha(g(y,X)) &= \frac{1}{\alpha} \int_0^{F^{-1}(\alpha)} g(y,x)\mathrm{d}F(x) \\ &= \frac{1}{\alpha}(p-z) \int_0^{F^{-1}(\alpha)} x\mathrm{d}F(x) - (c-z)y.\end{aligned} \tag{8.16}$$

$$\begin{aligned}\text{For } \alpha \geq F(y): \quad \mathrm{CVaR}_\alpha(g(y,X)) &= \frac{1}{\alpha}\left(\int_0^y g(y,x)\mathrm{d}F(x) + (p-c)y(\alpha - F(y))\right) \\ &= \frac{1}{\alpha}(p-z)\left(\int_0^y x\mathrm{d}F(x) - yF(y)\right) + (p-c)y.\end{aligned} \tag{8.17}$$

CVaR_α is monotonically increasing in α (see, e.g. (8.14)) and therefore

$$\mathrm{CVaR}_\alpha(g(y,X)) \leq \mathrm{CVaR}_1(g(y,X)) = E(g(y,X)).$$

Since the CVaR_α of a constant equals the constant the expected value $E(g(y,X))$ is preferred to $\mathrm{CVaR}_\alpha(g(y,X))$, and therefore for any $\alpha < 1$ the preference functional CVaR_α represents a risk-averse behavior; α is sometimes called "degree of risk aversion" (see, e.g., Chen et al. 2009).

Now the order quantity is derived that maximizes CVaR_α:

$$\max_y \mathrm{CVaR}_\alpha(g(y,X)). \tag{8.18}$$

Several authors consider the CVaR_α as objective function (see, e.g., Gotoh and Takano 2007; Gao et al. 2011). The CVaR measures the expected profit falling below a quantile level of the profit distribution. From (8.16) and (8.17), the solution of (8.18) can be derived:

$$\arg\max_y \text{CVaR}_\alpha(g(y,X)) = y^*_{\text{CVaR}}(\alpha) = F^{-1}\left(\alpha \frac{p-c}{p-z}\right). \tag{8.19}$$

As can be immediately seen, the CVaR order quantity (8.19) is only a fraction of the optimal order quantity y^* of the classical risk-neutral newsvendor given in (8.3), especially for small values of α. The higher the degree of risk aversion is, i.e., the smaller the α, the smaller is the order quantity (8.19).

Note that the optimal order quantity converges to $F^{-1}(\alpha)$ as $z \to c$. This implies that the decision maker will not order the maximal demand even if $P(g(y,X) \leq 0) = 0$. If the newsvendor can realize almost the same salvage value z for leftover products as the purchasing cost c, then the maximal profit $(p-c)F^{-1}(\alpha)$ is achieved. Of course the order quantity $y^*_{\text{CVaR}}(\alpha)$ is smaller than the order quantity $y^*(\alpha)$ of a VaR-newsvendor (see (8.10)) for all p, c, z.

The basic model with CVaR objective has been extended in several ways. Chen et al. (2009) consider the price-setting newsvendor with CVaR criterion. Xu (2010) analyzes this model with the possibility of emergency procurement, i.e., in a dual sourcing context where the newsvendor has a second ordering opportunity during the regular selling season if demand turns out to be larger than the first order quantity. Furthermore, the optimal solution is derived for positive shortage cost (Xu and Chen 2007). There are models for supply chain coordination using CVaR as objective function (see Yang et al. 2009). Also multi-product newsvendor models with CVaR objective have been investigated (see, e.g., Tomlin and Wang 2005; Choi et al. 2011).

8.5 Mean-CVaR Criteria

8.5.1 Convex Combination

Mean-deviation rules, e.g., the mean variance approach, are well known in portfolio theory. Gotoh and Takano (2007), Gao et al. (2011) and others consider mean deviation rules in the newsvendor context assuming that the deviation is measured by the CVaR of the profit.

The objective function is a convex combination of expected profit and CVaR

$$\gamma E(g(y,X)) + (1-\gamma)\text{CVaR}_\alpha(g(y,X)) \tag{8.20}$$

with $\gamma \in [0,1]$.

Note that for $\gamma = 1$, the classical newsvendor problem is given. For $\gamma < 1$, the objective function describes risk averse behavior, i.e., the expected profit $E(g(y,X))$ is preferred to $g(y,X)$. For $\gamma = 0$, we have the CVaR newsvendor of Sect. 8.4.

Of course, the degree of risk aversion increases the smaller γ and/or the smaller α; a small γ gives a high weight to the risk measure CVaR_α, a small α gives a high weight to high losses.

The solution of the above objective function is a special case of the approach in Sect. 8.5.2.

8.5.2 A General Mean-CVaR Criterion

In Jammernegg and Kischka (2007), a generalization of objective function (8.20) is provided. Consider first a profit variable Z with invertible distribution function F_Z. Let B denote a target profit and let $\alpha := F_Z(B)$. Then $E(Z|Z \le F_Z^{-1}(\alpha))$ and $E(Z|Z \ge F_Z^{-1}(\alpha))$ are conditional expected values of "bad" or "good" profits, respectively. Let $\lambda \in [0,1]$ be a weight of these expected profits (pessimism parameter).

Then the objective function is

$$\lambda E(Z|Z \le F_Z^{-1}(\alpha)) + (1-\lambda)E(Z|Z \ge F_Z^{-1}(\alpha)).$$

For the profit variable $g(y,X)$ in the newsvendor model, we use the generalized inverse of F_y^* introduced in Sect. 8.4.1 and replace the conditional expected values

$$\lambda \frac{1}{\alpha}\int_0^\alpha F_y^*(\beta)\mathrm{d}\beta + (1-\lambda)\frac{1}{1-\alpha}\int_\alpha^1 F_y^*(\beta)\mathrm{d}\beta. \tag{8.21}$$

This can be rewritten as (Jammernegg and Kischka 2007, p. 100)

$$\frac{1-\lambda}{1-\alpha}E(g(y,X)) + \frac{\lambda-\alpha}{1-\alpha}\text{CVaR}_\alpha(g(y,X)). \tag{8.22}$$

Note that for $\alpha > \lambda$, the objective function describes a risk-taking behavior, i.e., the random profit $g(y,X)$ is preferred to $E(g(y,X))$:

$$\begin{aligned}&\frac{1-\lambda}{1-\alpha}E(g(y,X)) + \frac{\lambda-\alpha}{1-\alpha}\text{CVaR}_\alpha(E(g(y,X)) \\ &\quad = E(g(y,X)) \le \frac{1-\lambda}{1-\alpha}E(g(y,X)) + \frac{\lambda-\alpha}{1-\alpha}\text{CVaR}_\alpha(g(y,X)).\end{aligned}$$

For $\alpha = \lambda$, $\alpha < \lambda$, risk-neutral and risk-averse behavior, respectively, prevails.

The objective functions (8.18) and (8.22) are consistent with dual utility theory. Dual utility theory as developed by Yaari (1987) is based on the idea that the probability of a bad result is judged differently from the same probability of a good result. Whereas in expected utility theory the monetary results are transformed with a utility function, in dual utility theory the probabilities of the monetary results are transformed. In Jammernegg and Kischka (2005), it is shown that for every pair (α,λ) with $0<\alpha<1, 0\le\lambda\le 1$ there exists a transformation of probabilities such that the objective function (8.22) is consistent with the axioms of dual utility theory.

Maximizing the objective functions (8.21) or (8.22), we get (Jammernegg and Kischka 2007, p. 101)

$$y^*(\alpha,\lambda)=\begin{cases} F^{-1}\left(\dfrac{p-c}{p-z}+\dfrac{\alpha-\lambda}{1-\lambda}\dfrac{c-z}{p-z}\right) & \text{for } \lambda\le\dfrac{p-c}{p-z} \\ F^{-1}\left(\dfrac{p-c}{p-z}\dfrac{\alpha}{\lambda}\right) & \phantom{\text{for }} \lambda\ge\dfrac{p-c}{p-z} \end{cases}. \tag{8.23}$$

Note that a risk-averse (risk-taking) decision maker orders less (more) than a risk-neutral decision maker. For a demand distribution with bounded support, the maximal demand is ordered as $z\to c$. This is in contrast with a related conclusion for the CVaR criterion in Sect. 8.4.

For $\lambda=1$, the solution for the objective function in Sect. 8.4 is given. For $\lambda\ge\alpha$, the objective function (8.20) with $\gamma=\frac{1-\lambda}{1-\alpha}$ is given. If $\lambda=\alpha$, i.e., $\gamma=1$, we have the special case of the classical risk-neutral newsvendor (8.2) and (8.3).

In the following, we compare the optimal order quantities and values of the objective functions. In Fig. 8.2, the graphs for the following objectives are shown: CVaR, VaR, expected profit, and risk taker with $\lambda=0$. The data are the same that have been used for the profit distributions in Fig. 8.1, especially $\alpha=0.5$.

As already mentioned before, Fig. 8.2 shows that the CVaR newsvendor is dominated by the VaR newsvendor also with respect to the expected profit if the respective optimal quantities are ordered. Therefore, CVaR_α is sometimes called a relatively conservative criterion. Figure 8.2 also shows that the expected profit curve is quite flat around its maximum; this is typical for many operations management models, think, e.g., of the economic order quantity model. Thus, a slight deviation from the optimum only leads to a small decrease of the expected profit; the optimal expected profit of the classical risk-neutral newsvendor is about 239 currency units ($y^*=160$ units) whereas for the most risk-taking newsvendor with $y^*(0.5,0)=230$ units the expected profit would be 220 currency units, a reduction in profit of about 8%. High order quantities result in high levels of customer service but also lead to a high probability to end up with a loss. Of course the opposite is true for low order quantities. We will come back to this trade-off between conflicting performance measures when dealing with constrained newsvendor models.

The following extensions of models using the mean-CVaR criterion consider the objective function (8.20). Like in the previous section there are multi-product newsvendor models (Choi et al. 2011) and models with positive shortage cost (Ahmed et al. 2007; Xu and Chen 2007). In a recent paper, the newsvendor not

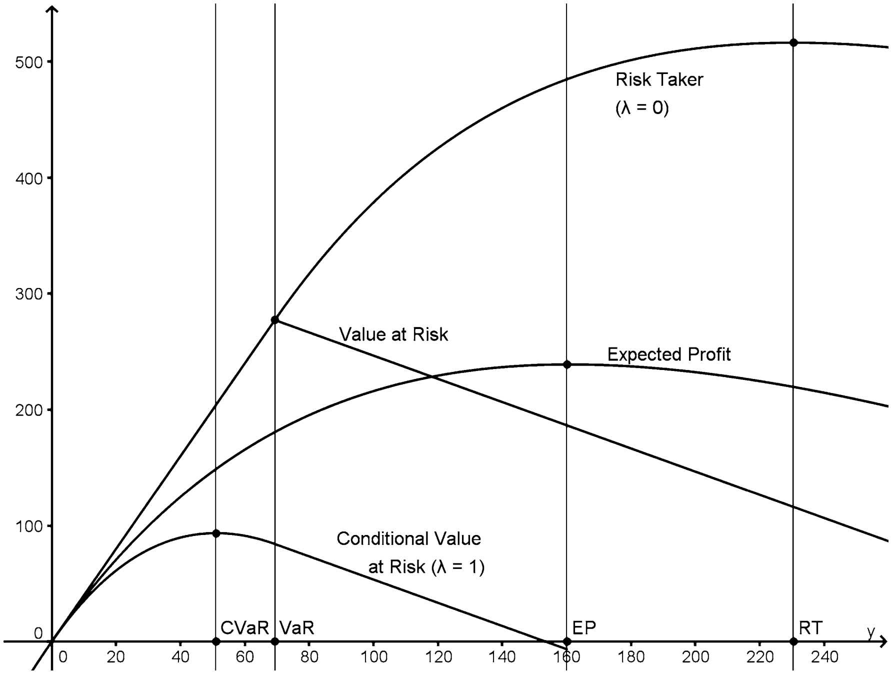


Fig. 8.2 CVaR-, VaR-, expected profit- and risk taking ($\lambda = 0$)-objectives and optimal order quantities

only decides the order quantity but also adopts a weather hedging strategy. Using the mean-CVaR criterion, the weather-derivative hedging can increase the order quantity and can improve the expected profit as well as the CVaR-profit (Gao et al. 2011).

8.6 Constraints

In this section, first we present a single-product newsvendor model with a VaR constraint, where the objective is to maximize the expected profit. Then we introduce a model with a loss constraint and a service constraint. The loss constraint specifies an upper bound for the probability resulting in loss, i.e., it is a special version of the VaR constraint where the target profit is equal to zero. In the service constraint, a lower bound for the cycle service level is prescribed.

8.6.1 VaR Constraint

Remember that for order quantity y, the maximum profit is $(p-c)y$ and the minimum profit is $(z-c)y$. Let B be some target profit (see Sect. 8.3.2) and let η be some probability with

$$P(g(y,X) \leq B) \leq \eta. \tag{8.24}$$

If the quantity y is ordered, the probability that the profit is not higher than the target profit B is at most η.

Consider the following newsvendor problem with a so-called VaR-constraint (Gan et al. 2005):

$$\begin{aligned} &\max_y E(g(y,X)) \\ &s.t. P(g(y,X) \leq B) = F_y(B) \leq \eta. \end{aligned} \tag{8.25}$$

Using (8.4) we can rewrite the constraint (8.24) as follows:

$$0 \leq y \leq \frac{F^{-1}(\eta)(p-z)-B}{c-z}. \tag{8.26}$$

Of course an admissible solution only exists if $B \leq F^{-1}(\eta)(p-z)$.

Remember that y^* denotes the solution of the classical newsvendor (see (8.3)). Then the solution of (8.25) is given by

$$y^*(\alpha,B) = \begin{cases} y^* & \text{for} \quad y^* \leq \dfrac{F^{-1}(\eta)(p-z)-B}{c-z} \\ \dfrac{F^{-1}(\eta)(p-z)-B}{c-z} & \phantom{\text{for}} \quad y^* \geq \dfrac{F^{-1}(\eta)(p-z)-B}{c-z} \end{cases}. \tag{8.27}$$

This result can be found in Gan et al. (2005), Özler et al. (2009), and Zhang et al. (2009). Gan et al. (2005) use this result to derive a coordinating contract between a risk-neutral supplier and a retailer with a VaR constraint. In Yang et al. (2007), the optimal order quantity is derived for the newsvendor model with positive shortage cost if the cost target is fixed or given by the expected cost. Özler et al. (2009) extend this model to a multi-product newsvendor problem with VaR constraint. In Zhang et al. (2009), instead of a VaR constraint a CVaR constraint is proposed. Furthermore, they use this framework for multi-period inventory models.

8.6.2 A Mean-CVaR Criterion With Service and Loss Constraints

In this section, we extend the approach of Sect. 8.6.1 in two ways (Jammernegg and Kischka 2011). First we use the general objective function (8.22); note that for $\alpha = \lambda$, the risk-neutral case of the above Sect. 8.6.1 is included.

Second we assume that the optimal order quantity is chosen according to some constraints given by performance measures. E.g., such constraints are also considered in Sethi et al. (2007) and Özler et al. (2009). Contrary to these papers, we simultaneously use internal and external performance measures. It is intuitively clear that these measures may collide and there is no admissible solution. In the following we discuss the set of admissible solutions and give the optimal order quantity under the constraints.

As internal performance measure we use the probability of loss, i.e., we consider the VaR-constraint (8.24) with $B = 0$:

$$P(g(y,X) \leq 0) \leq \eta. \tag{8.28}$$

From (8.26), it is clear that admissible solutions fulfilling (8.28) always exist.

As external performance measure, we use the cycle service level which is defined as the probability to fulfill demand, i.e., for order quantity y the cycle service level is given by $F(y)$:

$$y \geq F^{-1}(\delta). \tag{8.29}$$

Constraint (8.29) states that the cycle service level at least must be δ.

Combining (8.28) and (8.29), the set A of admissible solutions is given by

$$A = \left\{ y | F^{-1}(\delta) \leq y \leq \frac{F^{-1}(\eta)(p-z)}{c-z} \right\}. \tag{8.30}$$

Denoting the profit value of the product $pv = \frac{p-c}{p-z}$, we have the following condition for the existence of an admissible order quantity:

$$A \neq \emptyset \Leftrightarrow pv \geq 1 - \frac{F^{-1}(\eta)}{F^{-1}(\delta)}.$$

Thus, the higher the profit value of the product the more likely is the existence of an admissible solution. Moreover, a solution exists the larger is the prescribed acceptable probability of loss η and/or the smaller is the prescribed cycle service level δ.

There is no existence problem if either the internal or the external performance measure is considered. This can be easily seen if the probability of loss $\eta = 1$ and the cycle service level $\delta = 0$, respectively. For $\eta \geq \delta$, an admissible solutions always exists. Of course, this specification is not plausible from an economic point of view. For the relevant case, $\eta < \delta$ (8.30) represents the problem of considering internal and external performance measures simultaneously.

The optimal order quantity is the solution of the following constrained model (see (8.22), (8.30)):

$$\max_{y \in A} \frac{1-\lambda}{1-\alpha} E(g(y,X)) + \frac{\lambda - \alpha}{1-\alpha} \mathrm{CVaR}_\alpha(g(y,X)). \tag{8.31}$$

Table 8.1 Risk preferences for profit value pv, demand distribution F and target values δ and η (Jammernegg and Kischka 2011)

	$pv < 1 - \frac{F^{-1}(\eta)}{F^{-1}(pv)}$	$pv = 1 - \frac{F^{-1}(\eta)}{F^{-1}(pv)}$	$pv > 1 - \frac{F^{-1}(\eta)}{F^{-1}(pv)}$
$pv < \delta$	$A = \emptyset$	$A = \emptyset$	If $A \neq \emptyset$, risk-taking
$pv = \delta$	$A = \emptyset$	$A = \{F^{-1}(\delta)\}$, risk-neutral	$A \neq \emptyset$, risk-neutral or risk-taking
$pv > \delta$	If $A \neq \emptyset$, risk-averse	$A \neq \emptyset$, risk-neutral or risk-averse	$A \neq \emptyset$, all risk preferences

Solving (8.31) gives the optimal order quantity of an inventory manager who may have some special kind of risk preferences depending on the relation of α, λ (see Sect. 8.5.2) and simultaneously tries to fulfill some constraints concerning internal and external performance measures.

From Jammernegg and Kischka (2007), we know that the objective function (8.22) is a concave function of y. Therefore, if an admissible solution (see (8.30)) exists, there exists also an optimal solution of (8.31) which we denote by $\hat{y}(\alpha,\lambda)$. Moreover, from the concavity we can conclude

(1) $y^*(\alpha,\lambda) \leq F^{-1}(\delta) \Rightarrow \hat{y}(\alpha,\lambda) = F^{-1}(\delta)$

(2) $y^*(\alpha,\lambda) \geq \frac{F^{-1}(\eta)}{1-pv} \Rightarrow \widehat{y}(\alpha,\lambda) = \frac{F^{-1}(\eta)}{1-pv}$

(3) $F^{-1}(\delta) \leq y^*(\alpha,\lambda) \leq \frac{F^{-1}(\eta)}{1-pv} \Rightarrow \widehat{y}(\alpha,\lambda) = y^*(\alpha,\lambda)$.

E.g., if—for given α, λ—the optimal unrestricted solution $y^*(\alpha,\lambda)$ (see (8.23)) would exceed $\frac{F^{-1}(\eta)}{1-pv}$, then the solution of the restricted problem is the right corner of the set of admissible solutions (see (8.30)). Since $y^*(\alpha,\lambda)$ is increasing in α and decreasing in λ, we see that this situation is more probable for $\lambda < \alpha$, i.e., for a risk-taking inventory manager.

8.6.3 Deduction of Risk Parameters From Specified Target Values

Using the general mean-CVaR objective function (8.31) in Jammernegg and Kischka (2011), it is shown that every admissible solution $y \in A$ is optimal with respect to some (α,λ)-combination, i.e., some risk attitude. From the monotonicity properties of $y^*(\alpha,\lambda)$, we can deduce consistent risk attitudes from the prescribed performance measures. Thus, the newsvendor must not be able to specify the risk parameters. Instead the risk preferences can be derived from the target values for the probability of loss η and for the cycle service level δ. The results are summarized in Table 8.1.

A product is characterised by its profit value pv, its demand distribution F, the loss target and the service target. For these product characteristics Table 8.1 shows whether an admissible solution exists and in the positive case the associated risk preferences of the newsvendor are noted.

If the loss target is low and the service target is high with respect to the profit value and the demand distribution then no admissible order quantity exists; this is described by the upper left area in Table 8.1. Contrary, in any case admissible solutions exist if both target values η and δ are not very challenging for the specific product. In Table 8.1, this is represented by the lower right area. Especially for products with high profitability pv the decision maker can exploit any risk attitude.

If both constraints are fulfilled as equalities then the only admissible solution is the optimal order quantity (8.3) of the classical, riskneutral newsvendor which is represented in the centre of Table 8.1. Finally, in the boxes at end of the off-diagonal one constraint is dominating provided an admissible order quantity exists. The upper right corner characterizes a high service target for the product; here the newsvendor shows risktaking behavior. If the loss constraint is dominating the decision maker is a risk averter. This situation is shown in the lower left corner of Table 8.1

In Fig 8.3 we use the data of the examples in Figs 8.1 and 8.2 to illustrate the findings of Table 8.1 for fixed cycle service level $\delta = 0.6$ and different probabilities of loss η. Here, the vertical lines denote the corresponding boundaries of the admissible sets. Since the profit value of the product is $pv = 0.8$ the last row of Table 8.1 is relevant. If $\eta = 0.1$ and $\eta = 0.2$ then $pv < 1 - F^{-1}(\eta)/F^{-1}(pv)$ holds. As can be seen from Fig 8.3 for the low probability of loss $\eta = 0.1$ the loss constraint (upper bound) is smaller than the service constraint (lower bound); thus, no admissible solution exists. For $\eta = 0.2$ admissible solutions exist, but the newsvendor in any case is a risk averter (remember that the optimal risk-neutral order quantity y^* is 160 units). If in addition to the predetermined cycle service level $\delta = 0.6$ the probability of loss $\eta = 0.3$ is not very challenging, too, then $pv > 1 - F^{-1}(\eta)/F^{-1}(pv)$ holds, i.e. the existence of admissible solutions is guaranteed. From the lower right area in Table 8.1 we know that in this case the decision maker can exploit any risk preference. From Fig. 8.3 we see that the newsvendor is a risk averter if the chosen order quantity is from the interval [91.6, 160] in contrast, for an order quantity in the interval [160, 178.4] risk-taking behavior is expressed. As indicated before the classical risk-neutral newsvendor orders 160 units of the product.

The relationship of the product characteristics and the implied risk preferences may not match with the basic intentions of the responsible inventory manager. Then the findings from Table 8.1 also can be used to reposition the product. According to Fisher (1997) for a functional product it could be reasonable to increase the profit value by reducing the purchasing cost, e.g. by renegotiating the supply contract in place or to lower the loss constraint. Innovative products are characterized by high profitability. Thus, a high level of product availability is necessary to fulfil the entire demand in order to generate as much revenue as possible. Of course, in this case the service target should be increased. Especially for innovative products the demand distribution should be updated as soon as additional relevant information becomes available to make it less variable, e.g. by reducing its coefficient of variation.

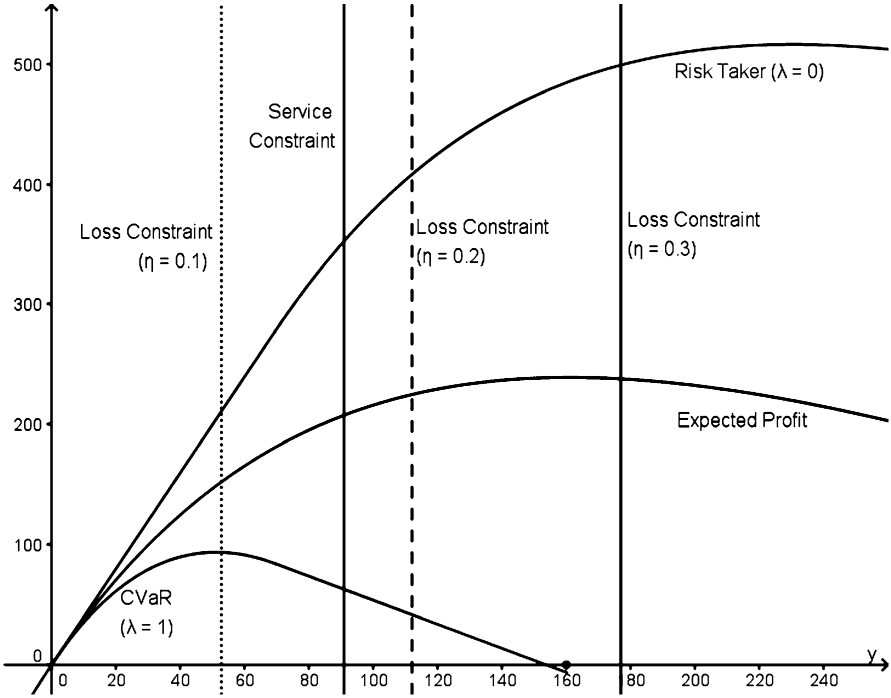


Fig. 8.3 CVaR-, expected profit- and risk taking ($\lambda = 0$)-objectives and admissible order quantities for probabilities of loss η = 0.1 (*dotted line*), $\eta = 0.2$ (*dashed line*) and $\eta = 0.3$ (*solid line*)

8.7 Spectral Risk Measures

So far we have considered approaches to formulate the risk preferences of a newsvendor by means of the risk measures VaR, CVaR, and mean-CVaR either as objective function or as constraint. Furthermore, we have presented some reasonable extensions of these basic models.

In this concluding section, we will consider some rather new developments in the theory of risk measures, which are also relevant for the newsvendor model. VaR and CVaR originated in the theory of finance. Because of lacking subadditivity and other deficiencies, the VaR risk measure is criticized and the focus is now on coherent risk measures like CVaR and the convex mean-CVaR measure.

Remember that the presented objective functions can be seen as negative risk measures; e.g., $-\mathrm{CVaR}_\alpha(g(y,X))$ is a coherent risk measure; this holds also for (8.2) and (8.20). A special subset of the coherent risk measures is the class of spectral risk measures (Acerbi 2002). It can be shown that every spectral risk measure ρ is of the form

$$\rho(Z) = -\int_0^1 F_Z^*(\beta)\varphi(\beta)\mathrm{d}\beta, \tag{8.32}$$

where F_Z^* is the generalized inverse of the random variable Z (see (8.13)) and φ denotes the so called risk spectrum, i.e.:

$$\begin{aligned} &\varphi : [0,1] \to \mathbb{R}_+ \\ &\int_0^1 \varphi(\beta)\mathrm{d}\beta = 1 \\ &\varphi \text{ is monotonically decreasing.} \end{aligned} \tag{8.33}$$

Conversely, every function φ fulfilling (8.33) defines a spectral risk measure (8.32).

The objective functions (8.2), (8.18), (8.19), and (8.20) all can be derived from a spectral risk measure; e.g., with

$$\varphi(\beta) = \begin{cases} \gamma + \frac{1}{\alpha}(1-\gamma) & \text{for } 0 \le \beta \le \alpha \\ \gamma & \phantom{\text{for }} \alpha < \beta \le 1 \end{cases}$$

we have for the corresponding risk measure ρ

$$-\rho(g(y,X)) = \gamma E(g(y,X)) + (1-\gamma)\mathrm{CVaR}_\alpha(g(y,X)),$$

which is the mean-CVaR measure (8.20) (Brandtner 2011).

In general, we have for the newsvendor problem with a spectral risk measure ρ

$$\begin{aligned} &\max_y -\rho(g(y,X)) \\ &\arg\max_y -\rho(g(y,X)) = y^*(\rho) = F^{-1}\left(\Phi^{-1}\left(\frac{p-c}{p-z}\right)\right), \end{aligned}$$

where Φ denotes the primitive of φ (Fichtinger 2010).

With the risk spectrum φ, the quantiles of Z, i.e., in the newsvendor context the quantiles of the profit distribution F_y can be weighted. With a special risk spectrum, e.g., the exponential risk spectrum, special kinds of risk aversion can be modeled (Fichtinger 2010; Brandtner 2011).

Spectral risk measures are also a subset of the set of convex risk measures (Föllmer and Schied 2002); first applications of convex risk measures to the newsvendor problem are given in Brandtner (2011).

8.8 Summary

In this chapter, we considered the single-product newsvendor model where the risk preferences of the decision maker were expressed by the risk measures VaR, CVaR, and (general) mean-CVaR. With the general mean-CVaR measures it is possible to describe not only risk-averse and risk-neutral but also risk-taking behavior. These risk measures were included in the newsvendor model as objective functions and as constraints. The basic intention of the paper are comparative analyses of the risk measures with respect to their impact on the distribution functions of profit as well as on the respective optimal order quantities and optimal profits.

For the presented basic models, we reviewed the literature and referred to extensions, e.g., multi-product models and models with price-dependent demand. Finally, we briefly described spectral risk measures where CVaR and mean-CVaR are special cases. A deeper analysis of these risk measures seems to be a promising stream for future research for newsvendor models with risk preferences.

References

Acerbi, C. (2002). Spectral measures of risk: a coherent representation of subjective risk aversion. *Journal of Banking and Finance, 26*, 1505–1526.

Acerbi, C., & Tasche, D. (2002). On the coherence of expected shortfall. *Journal of Banking and Finance, 26*, 1487–1503.

Ahmed, S., Cakmak, U., & Shapiro, A. (2007). Coherent risk measures in inventory problems. *European Journal of Operational Research, 182*, 226–238.

Artzner, P., Delbaen, F., Eber, J.M., & Heath, D. (1999). Coherent measures of risk. *Mathematical Finance, 9,* 203–228.

Brandtner, A. (2011). *Newsvendor-Modelle mit spektralen und konvexen Risikomaßen.* Working paper, Jena: Friedrich Schiller University.

Chen, Y., Xu, M., & Zhang, Z. (2009). A risk-averse newsvendor model with CVaR criterion. *Operations Research, 57,* 1040–1044.

Chiu, C.H., & Choi, T.M. (2010). Optimal pricing and stocking decisions for newsvendor problem with value-at-risk consideration. *IEEE Transactions on Systems Man, and Cybernetics–Part A: Systems and Humans, 40,* 1116–1119.

Choi, S., Ruszczinsky, A., & Zhao, Y. (2011). A multi-product risk-averse newsvendor with law-invariant coherent measures of risk. *Operations Research, 59,* 346–364.

Choi, T.M., & Chiu, C.H. (2012). Mean-downside-risk and mean-variance newsvendor models: implications for sustainable fashion retailers. *International Journal of Production Economics, 135*, 552–560.

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

Fichtinger, J. (2010). *The Single-period Inventory Model With Spectral Risk Measures.* WU Vienna: Doctoral Dissertation.

Fisher, M. (1997). *What is the right supply chain for your product?* Harvard Business Review, March-April, 83–93.

Föllmer, H., & Schied, A. (2002). Convex measures of risk and trading constraints. *Finance and Stochastics, 6,* 429–447.

Gan, X., Sethi, S., & Yan, H. (2005). Channel coordination with a risk-neutral supplier and a downside-risk-averse retailer. *Production and Operations Management, 14,* 80–89.

Gao, F., Chen, F.Y., & Chao, X. (2011). Joint optimal ordering and weather hedging decisions: mean-CVaR model. *Flexible Services and Manufacturing Journal, 23,* 1–25.

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

Jammernegg, W., & Kischka, P. (2005). A decision rule based on the conditional value at risk. *Jena: Friedrich Schiller University*.

Jammernegg, W., & Kischka, P. (2007). Risk-averse and risk-taking newsvendors: a conditional expected value approach. *Review of Managerial Science, 1,* 93–110.

Jammernegg, W., & Kischka, P. (2011) Risk preferences of a newsvendor with service and loss constraints. *International Journal of Production Economics.* doi:10.1016/j.ijpe.2011.10.017.

Lau, H. (1980). The newsboy problem under alternative optimization objectives. *Journal of the Operational Research Society, 31,* 525–535.

Li, J., Chen, J., & Wang, S. (2011). *Risk management of supply and cash flows in supply chains.* New York: Springer.

Özler, A., Tan, B., & Karaesmen, F. (2009). Multi-product newsvendor problem with value-at-risk considerations. *International Journal of Production Economics, 117,* 244–255.

Parlar, M., & Weng, Z. (2003). Balancing desirable but conflicting objectives in the newsvendor problem. *IIE Transactions, 35,* 131–145.

Qin, Y., Wang, R., Vakharia, A., Chen, Y., & Seref, M. (2011). The newsvendor problem: review and directions for future research. *European Journal of Operational Research, 213,* 361–374.

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

Sethi, S., Yan, H., Zhang, H., & Zhou, J. (2007). A supply chain with a service requirement for each market signal. *Production and Operations Management, 16,* 322–342.

Shi, C., & Chen, B. (2007). Pareto-optimal contracts for a supply chain with satisficing objectives. *Journal of the Operational Research Society, 58,* 751–759.

Tapiero, C. (2005). Value at risk and inventory control. *European Journal of Operational Research, 163,* 769–775.

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

Xu, M. (2010). A price-setting newsvendor model under CVaR decision criterion with emergency procurement. *International Journal of Systems Science and Systems Engineering, 19,* 85–104.

Xu, M., & Chen, F. (2007). Tradeoff between expected reward and conditional value-at-risk criterion in newsvendor models. *Proceedings of the 2007 IEEE International Conference on IEEM* (pp. 1553–1557) Singapore: IEEE Press.

Yaari, M. (1987). The dual theory of choice under risk. *Econometrica, 55,* 95–115.

Yang, L., Gao, C., Chen, K., & Li, J. (2007). Downside risk-aversion analysis for a single-stage newsvendor problem. *Wuhan University Journal of Natural Sciences, 12,* 198–202.

Yang, L., Xu, M., Yu, G., & Zhang, H. (2009). Supply chain coordination with CVaR criterion. *Asia-Pacific Journal of Operational Research, 26,* 135–160.

Yang S, Shi C, & Zhao X (2011) Optimal ordering and pricing decisions for a target oriented newsvendor. *Omega,* 39: 110–115.

Zhang, D., Xu, H., & Wu, Y. (2009). Single and multi-period optimal inventory control models with risk-averse constraints. *European Journal of Operational Research, 199,* 420–434.

Part II
Applications

Chapter 9
Production and Remanufacturing Strategies in a Closed-Loop Supply Chain: A Two-Period Newsvendor Problem

Marc Reimann and Gernot Lechner

Abstract Effective and efficient closed-loop supply chain processes can constitute a significant competitive edge for companies. However, the integration of forward and reverse processes poses some challenges both on the supply side—e.g., availability of remanufacturable products—and on the demand side, e.g., cannibalization between new and remanufactured products. In this paper a two-period newsvendor-type approach is presented. The model is used to characterize the optimal production and remanufacturing policies. The main emphasis is on studying supply side interactions, in particular, the link between production and sales of new products and the resulting subsequent supply of used products. Further, the issue of storing excess production is addressed. The relationship between inventory and remanufacturing decisions is quantified.

Keywords Closed-loop supply chain • Remanufacturing policies • Two-period newsvendor • Remanufacturable products • Inventory carryover

9.1 Introduction

Over the last few years the design of closed-loop supply chain operations has attracted increasing attention in several industries including prominently, e.g., automotive or consumer electronics (Guide et al. 2006; Olugu et al. 2011). The term closed-loop refers to the fact that forward processes and reverse production or logistics processes are dealt with in an integrated fashion. The reverse processes may include some or all of the following stages: product acquisition, quality grading, repair, remanufacturing, recycling, or disposal (Guide and Van Wassenhove 2009).

M. Reimann (✉) • G. Lechner
Institute of Production and Operations Management, University of Graz, Universitätsstrasse 15, A-8010 Graz, Austria
e-mail: marc.reimann@uni-graz.at; gernot.lechner@uni-graz.at

T.-M. Choi (ed.), *Handbook of Newsvendor Problems*, International Series in Operations Research & Management Science 176, DOI 10.1007/978-1-4614-3600-3_9,

In this paper, we study the production decisions of a firm for a single product with uncertain demand that can be supplied through manufacturing brand new products or remanufacturing returned cores from sales in previous periods. Specifically our model explicitly captures the fact that the supply of returned cores depends on manufacturing and supply decisions for brand new products in the past. Together with the demand uncertainty, this link gives rise to two interesting intertemporal phenomena. First, by increasing the supply in early periods the firm increases the availability of returned cores in the future. As remanufacturing returned cores is more efficient than producing brand new products this gives rise to reduced cost in later periods. However, this effect comes from an increase in first period cost and the trade-off has to be balanced. Second, given the demand uncertainty excess production in early periods not only increases the availability of returned cores in the future but also increases the overage in the early periods. Keeping this overage in stock for sale in future periods also influences the *demand* for remanufactured products. Thus, while the existence of both stocking excess production and remanufacturing returned cores should increase the incentives for excess production in early periods the two supply options are also to some extent substitutes and the main interesting question is under what conditions remanufacturing takes place at all and under what conditions it is the exclusive supply option chosen.

To answer these questions a stochastic single-product two-period model is formulated and analyzed. In a first step, the optimal remanufacturing policy is determined assuming that keeping inventories is not an option. This case can also be interpreted as a situation where the second period corresponds to the life cycle of a new generation of the product, e.g., smartphones, video game consoles. In such a setting, the stocked first period product could not be used toward the satisfaction of the second period demand (at least not without some rework). Then, this model is extended by allowing inventories and the main structural properties of this extended model are also analytically derived. Particularly, it is shown that when inventory cost is sufficiently small, no remanufacturing may take place even though returned cores are available. This setting can be observed, e.g., in the printer cartridge market.

The remainder of the paper is organized as follows. Section 9.2 presents related research and places the current model with respect to the existing scientific literature. Section 9.3 deals with the formal model definition and the theoretical results. The model is extended to include the possibility of inventories in Sect. 9.4. Section 9.5 concludes the paper with a short summary and an outlook on extensions of the presented work.

9.2 Related Work

We will split the discussion of the existing research in two parts. In the first part, we will cover the works on related topics in terms of product returns and remanufacturing, while in the second part we will focus on some of the recent works on two-period newsvendor models. Note that in neither part the review is meant to be exhaustive but rather should give a rough overview of some of the recent developments in these areas.

9.2.1 Product Returns and Remanufacturing

As mentioned in the introduction, there is a growing body of literature dealing with reverse logistics and closed-loop supply chains. There is a stream of literature dealing with game-theoretic models for analyzing competition and supply chain coordination in remanufacturing settings. A two-period model is used in Ferguson and Toktay (2006) to study possible competition on the remanufacturing market. They develop strategies on how to prevent a remanufacturing market entry of a competitor by collecting used items or collection and remanufacturing. Supply chains with different coordination mechanisms are studied in Bhattacharya et al. (2006) and Li et al. (2011). In Bhattacharya et al. (2006), the results show that the option to remanufacture increases order quantities and profits. It also leads to a higher service level for customers due to increased product availability. A higher cost difference between new and remanufactured products results in increased order quantities. In Li et al. (2011), a supply chain is studied under three different coordination settings (Stackelberg case, Nash case, inaccessible return information case). A retailer is confronted with stochastic demand and orders from a supplier who produces new and remanufactures used products. In the Nash case with simultaneous decisions concerning the manufacturing/order quantities, the quantities and profits are the lowest of the three cases.

Our model is most closely related to the stream of research dealing with (the quality of) product returns and the relationship between new and remanufactured products. The quality of product returns may be highly heterogeneous and there are several models that analyze the acquisition and/or grading process in remanufacturing. In a very simple setting without the possibility to grade goods before acquisition, in Galbreth and Blackburn (2006) the optimal acquisition quantity of used cores is shown to exceed the demand for remanufactured products. As remanufacturing cost depends on core quality, this strategy enables the remanufacturer to select only the high quality returns for actual remanufacturing and scrap the low quality returns. In Ferguson et al. (2009), the return of cores is given and a grading system is in place that categorizes these returns according to their quality. The decisions are how many cores of a particular quality class to remanufacture immediately, how many cores of a particular class to store for later remanufacturing, and how many cores to scrap. One of the main results is that the company always remanufactures the exact demand in each period. Moreover, the optimal strategies are intuitive in that it can be never optimal to store lower quality cores when higher quality ones are available. Analogously, it can be never optimal to scrap higher quality cores before all lower quality cores are scrapped. While these two (and related papers) provide insight into the acquisition and supply process for remanufacturing, they ignore the link with new products.

This link is addressed in Guide and Li (2010) where the influence of product and market characteristics on the potential cannibalization of new product sales by remanufactured products is derived through an empirical, field research. The main finding is that for the studied commercial product there seemed to be a potential for cannibalization, while for the consumer product the risk of cannibalization was small. The relationship between new and remanufactured products for demand satisfaction is also studied analytically in Ferrer and Swaminathan (2006, 2010), Inderfurth (2004), Kelle and Silver (1989), Li et al. (2010), Shi et al. (2011), Teunter and Flapper (2011), Wei et al. (2011), Zhou et al. (2011), and Zhou and Yu (2011).

In Li et al. (2010) and Wei et al. (2011), the focus is on the solution approach. In Li et al. (2010), a dynamic programming approach is developed for a multi-period production planning model including manufacturing, remanufacturing, and disposal decisions. The structure of the optimal control consists of two order-up-to levels for remanufacturing and manufacturing, respectively, and a threshold inventory level above which returned products are disposed of. Robust optimization is applied to an inventory-production planning model with remanufacturing and uncertain demand and supply of used products in Wei et al. (2011). Some numerical examples underpin the effectiveness of the approach and show the sensitivity of the key parameters concerning the solution. Particularly, holding and shortage costs are shown to have the strongest influence on the optimal production and remanufacturing decisions.

In Inderfurth (2004), a single-period, combined manufacturing/remanufacturing and inventory control problem is presented. The same capacity is used by manufacturing and remanufacturing processes, and stochastic demands for new and remanufactured products as well as stochastic returns of used products are considered. Downward substitution allows substitution of remanufactured items by new products but not vice versa. A main result is that the optimal solution deviates from the newsboy solution, particularly by a decreased inventory level of the remanufactured product and higher production of new products. In Shi et al. (2011), a stochastic model for deciding optimal production and remanufacturing quantities for a product portfolio is presented. Product demands are independent, for each product new and remanufactured units are perfect substitutes, the returns are of unknown quality and the amount of returned cores is a function of their acquisition price, which is also a decision variable. Even for the single-period case studied, the problem is hard to solve for larger sizes and so a Lagrangian relaxation-based approach is presented to obtain near-optimal solutions. The optimal strategy will always include (some) remanufacturing. The optimal acquisition and remanufacturing policies of a model with uncertain quality of returns are determined in Teunter and Flapper (2011). Considering stochastic demand, optimal newsboy-like solutions are derived and consequences of demand uncertainty are explored. Higher demands result in an increased optimal quantity of acquired cores and larger optimal remanufacturing-up-to levels. The value of quality information decreases when the demand uncertainty increases.

None of the above-mentioned papers addresses explicitly the link between previous sales and returns of used products. One of the first papers focusing on this link is Kelle and Silver (1989), where the case of planning reusable containers is considered. Returns are stochastic but depend on past sales, and due to loss sometimes new containers must be acquired. In Zhou et al. (2011), the different quality of returns is considered in a single-product, finite multi-period inventory model with stochastic demands. As a result, it is shown that the optimal policy for manufacturing, remanufacturing, and disposal has a simple form, represented by a sequence of constant control parameters. Numerical examples show significant cost reductions compared to two heuristics (pull policy with sorting, pull policy without sorting). In Zhou and Yu (2011), dynamic pricing allows to influence the uncertain supply of used products and random customer demands in a production–remanufacturing model. An exogenous selling price results in a simple policy, whereas considering the selling price as endogenous decision variable leads to an indecomposable state-dependent solution. In this case, the selling price decreases and the acquisition price increases with rising inventory of serviceable products but both decrease when the aggregate inventory level increases.

In Ferrer and Swaminathan (2006, 2010), the optimal supply quantities of new and remanufactured products are analyzed in similar settings as in our model. The case of perfect substitutability between new and remanufactured products is dealt with in a deterministic setting in Ferrer and Swaminathan (2006). Using a price-dependent demand function, it is shown that the possibility of remanufacturing induces the OEM to reduce early period prices for new products to stimulate sales and consequently provide a larger supply of returned cores for possible remanufacturing in later periods. Moreover, it is shown that in the given setting remanufacturing in later periods will always take place, either as an exclusive supply or jointly with the production of new products. In Ferrer and Swaminathan (2010), the model is extended to deal with imperfect substitutability of new and remanufactured products and the equilibrium prices and quantities of new and remanufactured products are derived under a simple demand competition setting. It is shown that in this setting, there may be market constellations where no remanufacturing takes place. As in Shi et al. (2011), the optimal strategy will always include (some) remanufacturing.

Our model differs from these approaches by the combined consideration of the following problem characteristics:

- Uncertain product demand
- Manufacturing and remanufacturing decision making, i.e., a closed-loop view
- Explicit link between sales in earlier periods and subsequent availability of returns for remanufacturing
- No market clearing and the possibility to store excess production for future use

9.2.2 Two-Period Newsvendor Models

In terms of our modeling approach, we follow a line of research utilizing variants and extensions of newsvendor-type models. Particularly, there is a recent interest in two-period newsvendor models for studying different types of flexibility for satisfying uncertain demands of product portfolios (see, e.g., accurate response in Cattani et al. 2008; Chung et al. 2008; Reimann 2011a; Zhang and Du 2009 and postponement strategies in Granot and Yin 2008 or Reimann 2011b).

All of these models extend the classical newsvendor model by allowing (some) production after the demand revelation, i.e., during the selling season. In Cattani et al. (2008), the optimal levels of preseason and selling season capacity are determined under the assumption that the selling season capacity can be allocated to the different products upon demand realization. Contrary to that, the selling season capacity has to be pre-allocated to the different products before the selling season in Chung et al. (2008). The two settings are systematically compared in Reimann (2011a) to study the value of flexibility induced by delayed capacity allocation. Slightly deviating from the setting in these three studies, in Zhang and Du (2009) the value of outsourcing for supplementing limited in-house capacity is studied in two settings. In one setting, both in-house production and outsourcing decisions take place prior to the selling season. In the other setting, outsourcing can be used as an emergency option upon demand realization. Price and order postponement strategies to enhance effectiveness are studied in Granot and Yin (2008). Price postponement refers to the possibility of setting the price in reaction to the demand information, while order postponement is similar to the above-mentioned strategies and corresponds to adjusting supply quantities in response to demand revelation. Finally, in Reimann (2011b) accurate response and postponement strategies are combined in that prior to the selling season some standard component is produced, while during the selling season this standard component is then customized to the observed product demands.

In all of these models, the first period is only a preparatory phase and there is no demand in this period. Moreover, the possibility to utilize selling season capacity, i.e. to make decisions under certainty greatly enhances profitability and reduces the preseason production under uncertainty. In contrast to that, our current model deals with demands in both periods. Moreover, the first period demand and consequently the first period sales will influence (some of) the second period supply, namely, the one for remanufacturing. Consequently, it may be optimal to increase first period supply. In the remainder of the paper, we will show under which conditions this is the case.

9.3 The Model

We consider a two-period model. The manufacturer offers new products in the first period and has the opportunity to offer new and remanufactured products in the second period. Remanufactured products are made from customer returns of period 1 sales. The core collection yield is denoted by γ, i.e., a fraction $0 \leq \gamma \leq 1$ of the units sold in period 1 are available for remanufacturing in period 2.

New and remanufactured products are perfect substitutes. The price in period $t = 1,2$ is given by p_t, while the production cost for new products is $c_t < p_t$. The cost savings associated with remanufacturing is $\delta \geq 0$, i.e., remanufacturing a collected core incurs cost of $c_2 - \delta$ in period 2. Demand D_t in both periods is uncertain with known probability density and cumulative distribution functions f_{D_t} and F_{D_t}, respectively. Throughout we will assume that the demand distributions are continuous and twice differentiable. Let d_t denote a demand realization in period t.

The expected sales quantity in the first period is a function of the first period production decision of new products q_1 and denoted by $S_{D_1}(q_1)$. It is given by $S_{D_1}(q_1) = \int_0^{q_1} u\, f_{D_1}(u)\, \mathrm{d}u + q_1\, [1 - F_{D_1}(q_1)]$. Given the core collection yield γ defined above, $\gamma S_{D_1}(q_1)$ units will be returned by customers and are available for remanufacturing in the second period. However, the manufacturer may decide not to remanufacture all of them, and its remanufacturing decision variable is given by $\hat{q}_2 \le \gamma S_{D_1}(q_1)$. Moreover, the manufacturer can decide to manufacture q_2 units of new products in period 2. Summarizing the total supply in period 2 is given by $\hat{q}_2 + q_2$ and the associated expected sales are $S_{D_2}(\hat{q}_2 + q_2)$.

For formulating our intertemporal optimization problem, let us assume that second period cash flows are discounted with a factor $0 \le \beta \le 1$. Then the objective of maximizing expected profits π is given by

$$\max_{q_1, q_2, \hat{q}_2} \pi = -c_1\, q_1 + p_1\, S_{D_1}(q_1) + \beta\, [-c_2\,(q_2 + \hat{q}_2) + \delta\, \hat{q}_2 + p_2\, S_{D_2}(q_2 + \hat{q}_2)]\,, \quad (9.1)$$

while the constraints are

$$\hat{q}_2 \le \gamma S_{D_1}(q_1), \quad (9.2)$$

$$q_1, q_2, \hat{q}_2 \ge 0. \quad (9.3)$$

Clearly, the constraints are convex and it is easy to verify that the objective function is concave. Consequently, the optimal solution is obtained by solving the set of KKT optimality conditions.

The structure of the optimal solution is summarized by the following result.

Proposition 1. *Depending on the shadow-price λ_R of the remanufacturing constraint (9.2), the three possible production scenarios are given by*

1. $\lambda_R = 0$ *(Exclusive, but limited remanufacturing in period 2):*
   $q_1 = F_{D_1}^{-1}(\frac{p_1 - c_1}{p_1})$ *and* $\hat{q}_2 = F_{D_2}^{-1}(\frac{p_2 - c_2 + \delta}{p_2})$ *and* $q_2 = 0$
2. $0 < \lambda_R < \beta\delta$ *(Exclusive, full remanufacturing in period 2):*
   $q_1 = F_{D_1}^{-1}(\frac{p_1 - c_1 + \gamma\lambda_R}{p_1 + \gamma\lambda_R})$ *and* $\hat{q}_2 = \gamma S_{D_1}(q_1) = F_{D_2}^{-1}(\frac{\beta(p_2 - c_2 + \delta) - \lambda_R}{\beta\, p_2})$ *and* $q_2 = 0$
3. $\lambda_R = \beta\delta$ *(Full remanufacturing and new production in period 2):*
   $q_1 = F_{D_1}^{-1}(\frac{p_1 - c_1 + \gamma\beta\delta}{p_1 + \gamma\beta\delta})$ *and* $\hat{q}_2 = \gamma S_{D_1}(q_1)$ *and* $q_2 = F_{D_2}^{-1}(\frac{p_2 - c_2}{p_2}) - \hat{q}_2$

Proof. All proofs are given in Appendix. □

The optimal scenario and the associated production and remanufacturing quantities can be obtained easily through line-search for the optimal λ_R. Note that in the first case, the first period decision corresponds exactly to the well-known unconstrained, single-period newsvendor solution. There is no new production in period 2 and total second period supply is through remanufacturing. The second period supply quantity corresponds again to the unconstrained, single-period newsvendor quantity, this time subject to the remanufacturing cost $(c_2 - \delta)$. Finally, in that scenario not all the collected cores are remanufactured.

In the second scenario, first period production exceeds the newsvendor quantity and all the returned cores are remanufactured and offered to the market in the second period. However, there is still no new production in period 2.

In the third scenario, there is again excess production in period 1 and all the returned cores are remanufactured and offered to the market in the second period. However, this supply is insufficient and consequently supplemented by new production. In this third scenario, the total supply in the second period corresponds exactly to the unconstrained, single-period newsvendor quantity under the manufacturing cost c_2, i.e., the quantity that would also be produced if remanufacturing was not possible.

Thus, an interesting observation is that while in scenarios 2 and 3 there is excess period 1 production, it goes along with a reduction of the optimal total supply in period 2 compared to scenario 1. The excess production in period 1 is used only to narrow the gap between the optimal unconstrained remanufacturing supply and the optimal supply associated with the more costly new production.

Note that this result is in line with the results in Ferrer and Swaminathan (2006) where for the deterministic, but price-dependent demand case lower prices (and consequently larger supply quantities) in early periods are found when remanufacturing possibilities exist. An interesting new result provided by our approach is the link between excess production in period 1 and the remanufacturing quantity decision. Whenever excess production in period 1 occurs, this implies that all the returned cores are used for remanufacturing.

Using the results given by Proposition 1, the three different scenarios can be characterized as a function of the core collection yield γ and the remanufacturing cost savings δ. This is shown in the following lemma.

Lemma 1. *The existence of excess period 1 production and new production in period 2 is characterized by the following conditions.*

(a) Excess period 1 production occurs whenever

$$\gamma S_{D_1}(q_1^{\mathrm{NV}}) < F_{D_2}^{-1}\left(\frac{p_2 - c_2 + \delta}{p_2}\right). \tag{9.4}$$

Assuming uniform demand distributions $D_1 \sim U(a_1, b_1)$ and $D_2 \sim U(a_2, b_2)$ excess period 1 production occurs whenever

$$\delta > \frac{p_2}{b_2 - a_2}\left[\gamma\left[E[D_1] - (b_1 - a_1)\frac{1}{2}\frac{c_1^2}{p_1^2}\right] - \left[E[D_2] + (b_2 - a_2)\left(\frac{1}{2} - \frac{c_2}{p_2}\right)\right]\right]. \tag{9.5}$$

(b) New production in period 2 occurs whenever

$$\gamma S_{D_1}(q_1) < F_{D_2}^{-1}\left(\frac{p_2 - c_2}{p_2}\right). \tag{9.6}$$

Assuming uniform demand distributions $D_1 \sim U(a_1, b_1)$ *and* $D_2 \sim U(a_2, b_2)$ *new production in period 2 occurs whenever*

– *If* $\gamma E[D_1] - \left[E[D_2] + (b_2 - a_2)\left(\frac{1}{2} - \frac{c_2}{p_2}\right)\right] > 0$

$$\delta < \sqrt{\frac{b_1 - a_1}{2\gamma\beta^2} \frac{c_1^2}{\gamma E[D_1] - \left[E[D_2] + (b_2 - a_2)\left(\frac{1}{2} - \frac{c_2}{p_2}\right)\right]}} - \frac{p_1}{\gamma\beta} \tag{9.7}$$

– *If* $\gamma E[D_1] - \left[E[D_2] + (b_2 - a_2)\left(\frac{1}{2} - \frac{c_2}{p_2}\right)\right] \leq 0$

$$\delta \geq 0.$$

Part (a) of Lemma 1 states that excess production in period 1 can only be optimal when the expected collection yield associated with the newsvendor quantity q_1^{NV} is insufficient to cover the optimal unconstrained remanufacturing supply in period 2. This once again highlights the link between excess production in period 1 and reduced supply in period 2 described above. According to the second part (b) of Lemma 1 new production in period 2 can only occur when the collection yield associated with the optimal first period production is insufficient to cover even the optimal supply quantity associated with new production. It is easy to verify that the second part of Lemma 1 is more limiting on γ than the first. Whenever γ is large, i.e. condition (9.4) is violated, remanufacturing does not influence the first period decision and there is no excess production in period 1. We are in scenario 1. When γ is at an intermediate level, i.e. (9.4) is satisfied but (9.6) is violated, we are in scenario 2, while scenario 3 occurs for small values of γ which satisfy (9.6).

This provides an interesting insight into the strategic relationship between the collection efficiency/effectiveness and the manufacturing decision. Investing in a better return rate (e.g., by increasing the price paid for collected cores, or by improving the logistics network for collecting cores) reduces the necessity to produce excessive units in early periods just to ensure sufficient supply of remanufacturable cores in later periods.

For the special case of uniform demand distributions the lemma also provides explicit bounds on the remanufacturing cost savings δ. We first observe that when expected second period demand or demand uncertainty (given by the gap $(b_2 - a_2)$) increases, both excess period 1 production and new production in period 2 are more likely. When the market expands, i.e., in the early phases of the life cycle, it pays to provide a base for capitalizing on the remanufacturing opportunities. In the extreme case when $\gamma E[D_1] - [E[D_2] + (b_2 - a_2)(\frac{1}{2} - \frac{c_2}{p_2})] \leq 0$ new production in period 2 occurs for all $\delta \geq 0$. Moreover, in that case the right-hand-side of condition (9.5) is smaller than zero and hence excess production occurs whenever $\delta \geq 0$. This can be easily understood. Observe first, that $\gamma E[D_1]$ is the maximum possible expected sales quantity when $q_1 = b_1$. Observe further, that $E[D_2] + (b_2 - a_2)\left(\frac{1}{2} - \frac{c_2}{p_2}\right)$ is the minimum second period production corresponding to the newsvendor quantity associated with the cost of new production. Both excess production in the first period and new production in the second period need to take place if the returned cores induced by the maximum quantity produced in the first period are insufficient to satisfy the minimum second period production quantity.

Finally we also observe the inverse relationship between the core collection yield γ and the remanufacturing cost savings δ. When γ increases, the savings associated with remanufacturing need to be larger to induce additional excess production. Analogously, δ needs to be smaller, i.e., the cost savings need to be smaller to induce new production in period 2 when the core collection efficiency increases, i.e., γ. In both cases, the increased returns from the same first period sales quantity reduce the necessity of costly actions like excess production in period 1 and new production in period 2.

9.3.1 Illustrative Example

Let us consider a small illustrative example to support the theoretical findings above. For reasons of simplicity, we will assume that market prices and production costs are constant and given by $p_1 = p_2 = p = 10$, $c_1 = c_2 = c = 8$. The discount factor $\beta = 0.9$. First period demand is given by a uniform distribution $D_1 \sim U(a_1, b_1)$ with $a_1 = 25$ and $b_1 = 75$.

The main aim of the numerical study is to analyze variations in the core collection yield γ, the remanufacturing cost savings δ, and the expected second period demand $E[D_2]$ on the supply strategy and expected profitability.

For comparison, we will consider a base case setting, where $D_2 \sim U(25, 75)$, $\gamma = 0.5$, and $\delta = 1$. Figures 9.1–9.3 all show the first period production of new products q_1, the second period production of new products q_2, and the second period remanufacturing of returned cores $\hat{q}_2$. Further, for ease of explanation the figures also show the optimal single-period newsvendor quantity q_1^{NV} and the available supply of returned cores $\gamma S_{D_1}(q_1)$.

Figure 9.1 focuses on variations of γ from $\gamma = 0.05$ to $\gamma = 1$ in steps of 0.05. We first observe that there is excess period 1 production, i.e., $q_1 > q_1^{\mathrm{NV}}$ for the entire range of γ. Thus, we are never in case 1 described above in Proposition 1. This can also be seen from the fact that $\hat{q}_2 = \gamma S_{D_1}(q_1)$, i.e., all the returned cores are remanufactured. Moreover, except for the situation $\gamma = 1$ there is new production $q_2 > 0$ implying that we are in case 3, while for $\gamma = 1$ we are in case 2.

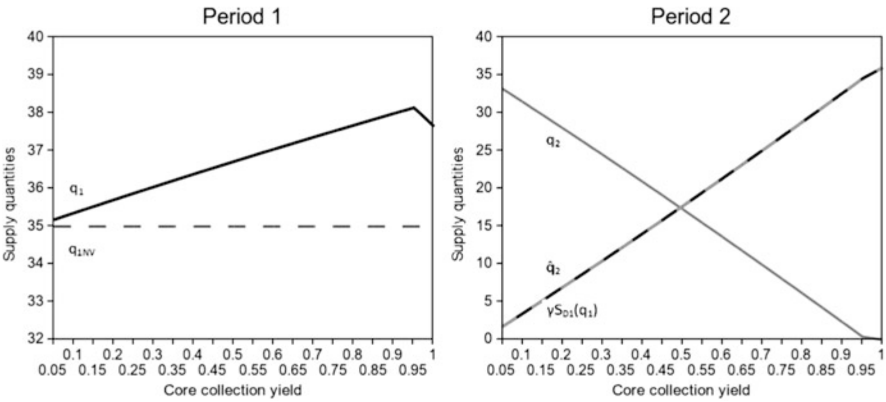


Fig. 9.1 Supply quantities under varying core collection yield γ

Looking at the general effect of γ, we observe some intuitive results. When the core collection yield increases, remanufacturing increases and second period new production falls. Expected profits increase almost linearly from 115.53 to 145.64 when γ goes from 0.05 to 1. However, compared to variations in δ and $E[D_2]$, the profit effect of variations in γ is quite moderate.

Finally, we observe an interesting and at first sight counterintuitive phenomenon. When γ increases q_1 first increases, and then drops again for $\gamma = 1$. When γ increases the same level of excess production yields a larger amount of returned cores. Thus, we would expect that *costly* excess production should go down. However, this is only true when γ increases from $\gamma = 0.95$ to $\gamma = 1$ in which case no more new production in period 2 takes place. For smaller values of γ, the core collection yield and excess first period production move in the same direction to enable an extra reduction of the more costly new production in period 2.

The impact of variations in δ between $\delta = 0$ and $\delta = c_2$ in steps of 0.5 are shown in Fig. 9.2. Here we observe very similar behavior as under variations of γ. When δ increases, and consequently remanufacturing gets less costly excess first period production increases, remanufacturing increases and new production in period 2 decreases. When $\delta = 0$, we obtain a special—and trivial—case. When remanufacturing yields no cost advantage, the two periods are decoupled and the optimal decision in both periods is to supply the unconstrained single-period newsvendor quantities q_1^{NV} and q_2^{NV}, respectively. As the base case is stationary both in terms of demand and cost structure, we observe $q_1 = q_2 = q_1^{NV} = q_2^{NV}$. Finally, we observe a steep expected profit increase of about 120.75% between the cases of $\delta = 0$ and $\delta = 8$. Specifically, the expected profit rises almost linearly from 114.00 when there is no cost difference between new and remanufactured products ($\delta = 0$), to 251.65, when the returned cores could directly be resold without any remanufacturing cost (i.e., $\delta = 8$).

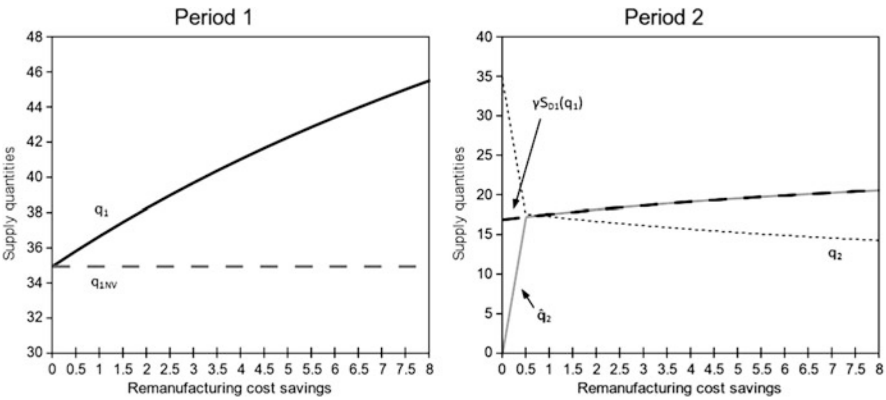


Fig. 9.2 Supply quantities under varying remanufacturing cost savings δ

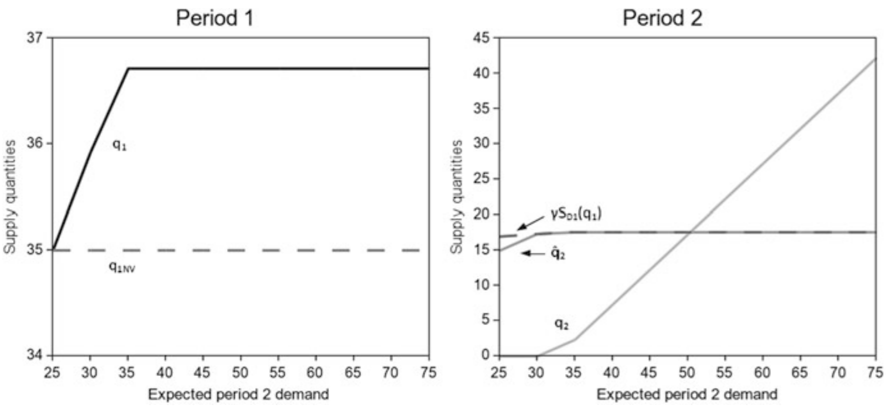


Fig. 9.3 Supply quantities under varying second period expected demand $E[D_2]$

Figure 9.3 shows the effects of different levels of second period demand. More precisely, we let $D_2 \sim U(a_2, b_2)$, where a_2 varies from 0 to 50 in steps of 5 and $b_2 = a_2 + 50$. Here we observe all three cases as described by Proposition 1. When second period demand is very small, i.e., $E[D_2] = 25$, we are in case 1. There is no new production in period 2, no excess production in period 1 and not all the returned cores are remanufactured. Supplying the unconstrained single-period newsvendor quantity in period 1 is sufficient to ensure enough returned cores for satisfying the very low second period demand. For $E[D_2] = 30$, there is still no new production in period 2, but all returned cores are remanufactured and first period production is already excessive. Thus, even on declining markets, excessive production and remanufacturing may be optimal when the decrease in demand from one period to the next is not too severe. Further, when $E[D_2] \geq 35$ full remanufacturing is supplemented by new production in period 2 to ensure sufficient second period supply. Finally, with increasing second period demand expected profits increase roughly linearly from 80.25 in the case of $E[D_2] = 25$ to 174.61 when $E[D_2] = 75$.

9.4 Inventory Carryover

In the previous section we have assumed that second period supply can only come from one of two sources, namely, production of new products or remanufacturing of returned cores in period 2. However, due to the demand uncertainty and pronounced through excess production in period 1, there may be unsold units of the product at the end of period 1. These could be carried over and used for demand satisfaction in period 2. This may not be a reasonable setting when the two periods correspond, e.g., to life cycles of two generations of a product. However, when the life cycle of a single product generation is considered inventories can play an important role. The interesting question is how these inventories will influence the optimal remanufacturing decision.

To answer this question, we will extend our model to deal with the possibility of inventory carryover. Given first period production q_1 and sales $S_{D_1}(q_1)$, there is an expected inventory level $I_{D_1}(q_1)$ of unsold units at the end of period 1 which is given by $I_{D_1}(q_1) = q_1 - S_{D_1}(q_1)$. The manufacturer may be able to keep (some of) these excess units for sale as new products in period 2. The associated per-unit holding cost is given by h. To avoid the case where holding can never be optimal (which would take us back to our previous model), we assume that $h < \beta c_2$. Further, to avoid the case where producing and holding units beyond the maximum demand in period 1 can be optimal we assume that $c_1 + h > \beta c_2$.

Holding excess production from period 1 is at the discretion of the manufacturer and the decision variable denoting the amount held is I_1. Clearly, $I_1 \leq I_{D_1}(q_1)$. The total supply in period 2 is given by $I_1 + \hat{q}_2 + q_2$ and the associated expected sales are $S_{D_2}(I_1 + \hat{q}_2 + q_2)$.

The modified model is given by

$$\max_{q_1,q_2,\hat{q}_2,I_1} \pi = -c_1 q_1 + p_1 S_{D_1}(q_1) - h I_1$$
$$+\beta\left[-c_2(q_2+\hat{q}_2) + \delta\hat{q}_2 + p_2 S_{D_2}(I_1 + q_2 + \hat{q}_2)\right], \tag{9.8}$$

while the constraints are

$$\hat{q}_2 \leq \gamma S_{D_1}(q_1), \tag{9.9}$$

$$I_1 \leq I_{D_1}(q_1), \tag{9.10}$$

$$q_1, q_2, \hat{q}_2, I_1 \geq 0. \tag{9.11}$$

As in the previous section, the model can be easily shown to be well behaved. The relationship between inventory and remanufacturing as supply sources is summarized in the following result.

Proposition 2. *There is a threshold holding cost $h^* = \beta(c_2 - \delta)$ such that, if $h > h^*$, remanufacturing is the primary supply for satisfying second period demand, while inventory as a secondary and new production as a third supply are only used to fill demand if necessary.*

Specifically, depending on the shadow-price λ_R of the remanufacturing constraint the possible production scenarios are given by

1. $\lambda_R = 0$ *(Exclusive, but limited remanufacturing in period 2):*
$q_1 = F_{D_1}^{-1}(\frac{p_1 - c_1}{p_1})$ *and* $\hat{q}_2 = F_{D_2}^{-1}(\frac{p_2 - c_2 + \delta}{p_2})$ *and* $I_1 = 0$ *and* $q_2 = 0$
2. $0 < \lambda_R < h - \beta(c_2 - \delta)$ *(Exclusive, full remanufacturing in period 2):*
$q_1 = F_{D_1}^{-1}(\frac{p_1 - c_1 + \gamma\lambda_R}{p_1 + \gamma\lambda_R})$ *and* $\hat{q}_2 = \gamma S_{D_1}(q_1) = F_{D_2}^{-1}(\frac{\beta(p_2 - c_2 + \delta) - \lambda_R}{\beta p_2})$ *and* $I_1 = 0$ *and* $q_2 = 0$
3. $\lambda_R = h - \beta(c_2 - \delta)$ *(Full remanufacturing and limited use of inventory in period 2):*
$q_1 = F_{D_1}^{-1}(\frac{p_1 - c_1 + \gamma(h - \beta(c_2 - \delta))}{p_1 + \gamma(h - \beta(c_2 - \delta))})$ *and* $\hat{q}_2 = \gamma S_{D_1}(q_1)$ *and* $I_1 = F_{D_2}^{-1}(\frac{\beta p_2 - h}{\beta p_2}) - \hat{q}_2$ *and* $q_2 = 0$
4. $h - \beta(c_2 - \delta) < \lambda_R < \beta\delta$ *(Full remanufacturing and full use of inventory in period 2):*
$q_1 = F_{D_1}^{-1}(\frac{p_1 - c_1 + \gamma\lambda_R}{p_1 + \gamma\lambda_R - [\beta(c_2 - \delta) - h + \lambda_R]})$ *and* $\hat{q}_2 = \gamma S_{D_1}(q_1)$ *and* $I_1 = I_{D_1}(q_1)$ *and* $q_2 = 0$
5. $\lambda_R = \beta\delta$ *(Full remanufacturing, full use of inventory, and new production in period 2):*
$q_1 = F_{D_1}^{-1}(\frac{p_1 - c_1 + \gamma\beta\delta}{p_1 + \gamma\beta\delta - (\beta c_2 - h)})$ *and* $\hat{q}_2 = \gamma S_{D_1}(q_1)$ *and* $I_1 = I_{D_1}(q_1)$ *and* $q_2 = F_{D_2}^{-1}(\frac{p_2 - c_2}{p_2}) - \hat{q}_2 - I_1$.

If $h \leq h^*$ inventory is the primary supply for satisfying second period demand, while remanufacturing as a secondary and new production as a third supply are only used to fill demand if necessary.

Specifically, depending on the shadow-price λ_I of the inventory constraint the possible production scenarios are given by

6. $\lambda_I = 0$ *(Exclusive, but limited use of inventory in period 2):*
$q_1 = F_{D_1}^{-1}(\frac{p_1 - c_1}{p_1})$ *and* $I_1 = F_{D_2}^{-1}(\frac{\beta p_2 - h}{\beta p_2})$ *and* $\hat{q}_2 = 0$ *and* $q_2 = 0$
7. $0 < \lambda_I < \beta(c_2 - \delta) - h$ *(Exclusive, full use of inventory in period 2):*
$q_1 = F_{D_1}^{-1}(\frac{p_1 - c_1}{p_1 - \lambda_I})$ *and* $I_1 = I_{D_1}(q_1) = F_{D_2}^{-1}(\frac{\beta p_2 - h - \lambda_I}{\beta p_2})$ *and* $\hat{q}_2 = 0$ *and* $q_2 = 0$
8. $\lambda_I = \beta(c_2 - \delta) - h$ *(Full use of inventory and limited remanufacturing in period 2):*
$q_1 = F_{D_1}^{-1}(\frac{p_1 - c_1}{p_1 - [\beta(c_2 - \delta) - h]})$ *and* $I_1 = I_{D_1}(q_1)$ *and* $\hat{q}_2 = F_{D_2}^{-1}(\frac{p_2 - c_2 + \delta}{p_2}) - I_1$ *and* $q_2 = 0$
9. $\beta(c_2 - \delta) - h < \lambda_I < \beta c_2 - h$ *(Full use of inventory and full remanufacturing in period 2):*
$q_1 = F_{D_1}^{-1}(\frac{p_1 - c_1 + \gamma[h - \beta(c_2 - \delta) + \lambda_I]}{p_1 + \gamma[h - \beta(c_2 - \delta) + \lambda_I] - \lambda_I})$ *and* $I_1 = I_{D_1}(q_1)$ *and* $\hat{q}_2 = \gamma S_{D_1}(q_1)$ *and* $q_2 = 0$

10. *$\lambda_I = \beta c_2 - h$ (Full use of inventory, full remanufacturing, and new production in period 2):*
    $q_1 = F_{D_1}^{-1}(\frac{p_1 - c_1 + \gamma\beta\delta}{p_1 + \gamma\beta\delta - (\beta c_2 - h)})$ *and* $I_1 = I_{D_1}(q_1)$ *and* $\hat{q}_2 = \gamma S_{D_1}(q_1)$ *and* $q_2 = F_{D_2}^{-1}(\frac{p_2 - c_2}{p_2}) - \hat{q}_2 - I_1$.

The results in Proposition 2 can be summarized as follows. The possibility of keeping inventories never increases the first period excess production and never decreases total second period supply. Both effects can be easily understood. First period excess production not only leads to increased sales but also leads to increased overages. While increased sales induce increased core collection for possible remanufacturing, overages can be used directly as a second period supply. Thus, for the same first period production, the total available units for second period supply are larger when inventories are kept. Consequently, the same second period supply can be achieved with smaller first period excess production.

Further, remanufacturing may not take place at all. The conditions for this as well as excess period 1 production and new production in period 2 are given in the following lemma.

Lemma 2. *The existence of excess period 1 production, no remanufacturing, and new production in period 2 is characterized by the following conditions.*

(a) Excess period 1 production occurs whenever

– *If $h > h^*$*

$$\gamma S_{D_1}\left(q_1^{\mathrm{NV}}\right) < F_{D_2}^{-1}\left(\frac{p_2 - c_2 + \delta}{p_2}\right) \tag{9.12}$$

– *If $h \leq h^*$*

$$I_{D_1}\left(q_1^{\mathrm{NV}}\right) < F_{D_2}^{-1}\left(\frac{\beta p_2 - h}{\beta p_2}\right). \tag{9.13}$$

(b) No remanufacturing takes place whenever

$$h \leq h^* \text{ and } I_{D_1}(q_1) > F_{D_2}^{-1}\left(\frac{p_2 - c_2 + \delta}{p_2}\right). \tag{9.14}$$

(c) New production in period 2 takes place whenever

$$\gamma S_{D_1}(q_1) + I_{D_1}(q_1) < F_{D_2}^{-1}\left(\frac{p_2 - c_2}{p_2}\right). \tag{9.15}$$

This is independent of the level of holding costs h.

Looking first at these results for $h > h^*$ we observe that the condition for excess period 1 production is identical to the one presented in Lemma 1. Moreover, remanufacturing will always take place as it is the primary supply option in period 2. Finally, the condition for new production in period 2 is more strict than the one given in Lemma 1. When inventories are possible, new production in period 2 can only occur when the sum of the collected cores and the inventory level associated with the optimal first period production is insufficient to cover the optimal supply quantity associated with new production. As the most costly new production only occurs when both inventory and remanufacturing are fully utilized, this condition also applies for $h \leq h^*$ and is actually independent of the level of holding cost.

Turning now to the results for $h \leq h^*$, the decision on excess period 1 production depends on whether or not the actual level of inventory from the unconstrained single-period newsvendor quantity in period 1 is sufficient to optimally supply the second period. Remanufacturing may not take place when inventory cost is low and the level of inventory exceeds the optimal second period supply associated with remanufacturing. From condition (9.14), we can see directly that remanufacturing is more likely when the associated cost savings δ or the second period price p_2 increase, or the second period cost c_2 decreases. Also an increase in second period demand will lead to more remanufacturing. The effect of first period characteristics is more implicit through the value of $I_{D_1}(q_1)$. $I_{D_1}(q_1)$ increases when the first period demand variance or service level (i.e., the optimal first period supply) increases. The latter increases when first period price increases or cost decreases. Consequently, in these cases it is more likely that no remanufacturing will occur.

Assuming uniform demand distributions, it is possible to derive closed-form expressions for the conditions on excess period 1 production, new production in period 2, and the occurrence of remanufacturing similar to those presented in Lemma 1. However, these expressions are more complex (third degree polynomials) and yield little explicit insight. Thus, we will now again turn to the numerical analysis of our illustrative example.

9.4.1 Illustrative Example

To show the effects of the possibility to keep inventories of unsold new production in period 1, we will return to our numerical setting from Sect. 9.3.1 and extend it by varying the holding cost h between 0 and 7.2 in steps of 1.2.

The results are shown in Fig. 9.4 for the basecase $\gamma = 0.5$, $\delta = 1$, and $E[D_2] = 50$. Figure 9.4 provides the same information as Figs. 9.1–9.3 and additionally shows the available and utilized inventory at the end of period 1 $I_{D_1}(q_1)$ and I_1, respectively. Let us first consider the settings that apply under different levels of h. In the numerical example $h^* = \beta(c_2 - \delta) = 6.3$. Thus, only for very high holding cost, in our case $h = 7.2$, remanufacturing is the primary second period supply, while in all the other cases inventories are the primary second period supply. Second, the case $h = 7.2$ implies that the manufacturer is indifferent between new production in period 2 and keeping unsold period 1 production in inventory for sale in period 2. Thus, while the figure shows some level of utilized inventory for $h = 7.2$, the expected profit, optimal first period production q_1 and optimal level of remanufacturing $\hat{q}_2$ are identical to the results from the base case without inventory. Keeping this in mind, we observe some interesting effects. As expected,

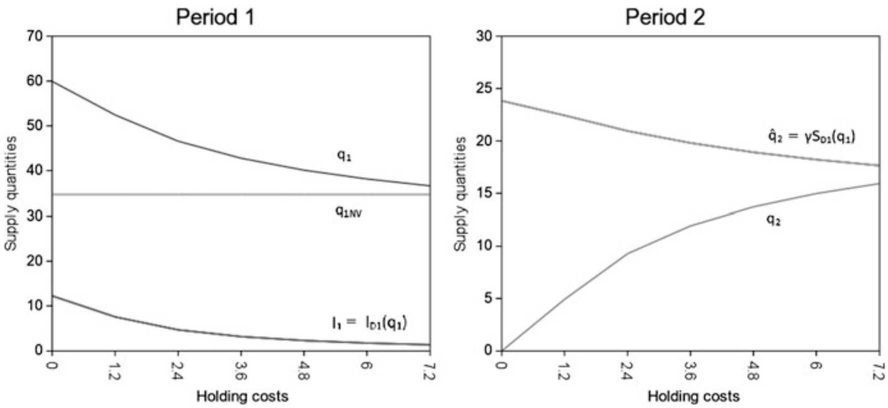


Fig. 9.4 Supply quantities under varying holding cost h

reducing the holding cost increases the expected profit, in particular, from 129.61 when $h = 7.2$ to 161.07 in the case of $h = 0$. Thus, whenever keeping inventories is possible and economically viable (i.e., the manufacturer is not indifferent to its utilization) the manufacturer increases its efficiency. The interesting fact is that this is achieved with increasing excessive new production in period 1. Thus, the manufacturer is willing to give up more and more period 1 profit in exchange for reduced cost in period 2. Finally, looking at period 2 supply one observes that remanufacturing and inventory complement each other in substituting the more costly new production in period 2. Thus, with decreasing holding cost h both the level of remanufacturing and the utilized inventory increase continuously.

To see whether this observation from the base case holds true in general and to study the effects of the possibility to store unsold period 1 production more thoroughly let us perform similar sensitivity analysis as in Sect. 9.3.1. To keep the unit gains from inventory at a comparable level with the remanufacturing cost savings we set $h = 1.2$ for the following experiments.

Tables 9.1–9.3 show the results of varying γ, δ, and $E[D_2]$ under both the model without inventories and the model with inventories. Note that for the model without inventories, the tables reproduce the results from Figs. 9.1–9.3, respectively.

Concerning the relationship between inventories and remanufacturing, we first observe from Table 9.1 that the result discussed above does not hold in general. When γ is small, I_1 and $\hat{q}_2$ both increase with increasing core collection yield. This is in line with the findings from above. However, when γ is large inventories start to drop as γ increases further, while $\hat{q}_2$ keeps increasing. This happens when $q_2 = 0$. In that case—as described above—q_1 starts to fall. While this fall translates directly into a fall in available and utilized inventory, the increasing core collection yield outweighs the reduction in q_1 and $\hat{q}_2$ increases.

Comparing in more detail the cases with and without inventories, we find that profits increase in both cases when γ increases. More interestingly, the gap between

Table 9.1 Supply quantities and expected profits with and without inventories under varying core collection yield γ and holding cost $h = 1.2$

		Core collection yield γ									
		0.05	0.1	0.15	0.2	0.25	0.3	0.35	0.4	0.45	0.5
π	w/o inv.	115.53	117.07	118.62	120.17	121.73	123.29	124.86	126.44	128.02	129.61
	w/ inv.	130.97	132.95	134.93	136.92	138.92	140.92	142.92	144.94	146.95	148.97
q_1	w/o inv.	35.18	35.36	35.53	35.71	35.88	36.05	36.22	36.39	36.56	36.72
	w/ inv.	50.28	50.55	50.82	51.08	51.34	51.58	51.83	52.07	52.30	52.53
q_2	w/o inv.	33.29	31.57	29.84	28.09	26.33	24.55	22.76	20.96	19.15	17.33
	w/ inv.	26.42	24.07	21.71	19.35	16.97	14.58	12.18	9.78	7.37	4.95
$\hat{q}_2$	w/o inv.	1.71	3.43	5.16	6.91	8.67	10.45	12.24	14.04	15.85	17.67
	w/ inv.	2.19	4.40	6.62	8.86	11.10	13.35	15.62	17.90	20.18	22.48
I_1	w/o inv.	–	–	–	–	–	–	–	–	–	–
	w/ inv.	6.39	6.53	6.67	6.80	6.94	7.07	7.20	7.33	7.45	7.58
		Core collection yield γ									
		0.55	0.6	0.65	0.7	0.75	0.8	0.85	0.9	0.95	1
π	w/o inv.	131.20	132.80	134.41	136.02	137.63	139.25	140.88	142.51	144.15	145.64
	w/ inv.	151.00	153.03	154.85	156.25	157.24	157.86	158.14	158.16	158.16	158.16
q_1	w/o inv.	36.89	37.05	37.21	37.37	37.53	37.69	37.84	38.00	38.15	37.65
	w/ inv.	52.75	52.97	51.57	50.12	48.71	47.30	45.92	45.41	45.41	45.41
q_2	w/o inv.	15.49	13.64	11.78	9.91	8.03	6.14	4.24	2.32	0.40	0.00
	w/ inv.	2.52	0.08	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
$\hat{q}_2$	w/o inv.	19.51	21.36	23.22	25.09	26.97	28.86	30.76	32.68	34.60	36.05
	w/ inv.	24.78	27.09	28.93	30.67	32.31	33.86	35.31	35.83	35.83	35.83
I_1	w/o inv.	–	–	–	–	–	–	–	–	–	–
	w/ inv.	7.70	7.82	7.06	6.31	5.62	4.97	4.38	4.17	4.17	4.17

the expected profits with and without inventory increase first and then decreases. Thus, the gain from the possibility to hold inventories is largest when γ is around 0.5. Further, first period excess production is much larger when inventories are possible. This is clear as excess production increases expected sales and expected inventory at the same time. Another interesting finding is that for small levels of γ, remanufacturing is larger with inventories, while for large values of γ the opposite is true. This can be understood by the fact that for $h = 1.2$ inventories are the primary supply and their utilization reduces the need for remanufacturing. Particularly, for large γ there are no longer all returned cores used for remanufacturing when inventories are possible.

Looking at Table 9.2, we observe some more interesting facts. The positive effect of increasing remanufacturing cost savings δ is actually pronounced by the possibility of keeping inventories, as indicated by the widening gaps between the expected profits with and without inventories. Concerning the relationship between I_1 and $\hat{q}_2$ we observe the same effect as already discussed for variations of h. When δ increases, both remanufacturing and utilized inventories increase throughout.

Finally, looking at Table 9.3 we find that expected second period demand has the strongest impact on new production in period 2 q_2. While I_1 and $\hat{q}_2$ also increase when the second period market size increases, once expected second period demand exceeds expected first period demand the additional demand is met exclusively through new production. Thus, in that case the cost savings associated with keeping

Table 9.2 Supply quantities and expected profits with and without inventories under varying remanufacturing cost savings δ and holding cost $h = 1.2$

		Remanufacturing cost savings δ							
		0.5	1	1.5	2	2.5	3	3.55	4
π	w/o inv.	121.73	129.61	137.63	145.79	154.07	162.47	170.98	179.59
	w/ inv.	138.92	148.97	159.14	169.41	179.76	190.19	200.69	211.24
q_1	w/o inv.	35.88	36.72	37.53	38.30	39.05	39.76	40.44	41.10
	w/ inv.	51.33	52.53	53.62	54.59	55.49	56.31	57.07	57.76
q_2	w/o inv.	17.65	17.33	17.02	16.73	16.46	16.21	15.97	15.75
	w/ inv.	5.87	4.95	4.10	3.33	2.61	1.94	1.32	0.75
$\hat{q}_2$	w/o inv.	17.35	17.67	17.98	18.27	18.54	18.79	19.03	19.25
	w/ inv.	22.20	22.48	22.71	22.92	23.10	23.25	23.39	23.51
I_1	w/o inv.	–	–	–	–	–	–	–	–
	w/ inv.	6.93	7.58	8.19	8.76	9.29	9.80	10.28	10.73
		Remanufacturing cost savings δ							
		4.5	5	5.5	6	6.5	7	7.5	8
π	w/o inv.	188.31	197.11	206.01	214.98	224.04	233.17	242.38	251.65
	w/ inv.	221.85	232.50	243.17	253.86	264.57	275.30	286.05	296.81
q_1	w/o inv.	41.74	42.35	42.94	43.50	44.05	44.58	45.09	45.59
	w/ inv.	58.40	58.83	59.12	59.39	59.65	59.90	60.15	60.38
q_2	w/o inv.	15.53	15.33	15.14	14.96	14.79	14.63	14.47	14.33
	w/ inv.	0.22	0.00	0.00	0.00	0.00	0.00	0.00	0.00
$\hat{q}_2$	w/o inv.	19.47	19.67	19.86	20.04	20.21	20.37	20.53	20.67
	w/ inv.	23.62	23.69	23.74	23.78	23.82	23.86	23.90	23.93
I_1	w/o inv.	–	–	–	–	–	–	–	–
	w/ inv.	11.16	11.45	11.64	11.82	12.01	12.18	12.35	12.51

inventories and remanufacturing do not outweigh the first period profit decrease due to increased excess production.

When looking at $\hat{q}_2$, we observe a similar effect as discussed above for variations of γ. When expected second period demand is small, $\hat{q}_2$ is smaller in the model with inventories than in the model without inventories. This effect is reversed when the expected second period demand increases—and in our numerical case exceeds $E[D_2] = 35$. The driver for this is again that for small expected second period demand not all of the returned cores are used. Specifically, we observe, e.g., for $E[D_2] = 25$ that the optimal second period supply in both models is 15. Without inventories this comes exactly from remanufacturing, while in the model with inventories this supply is split between inventories and remanufacturing. The most interesting observation is that even though the market is clearly declining—from $E[D_1] = 50$ to $E[D_2] = 25$—there is significant increased excess production in period 1 when inventories are possible. Yet this still leads to a considerable increase in expected profits.

Table 9.3 Supply quantities and expected profits with and without inventories under varying expected second period demand $E[D_2]$ and holding cost $h = 1.2$

		Expected second period demand $E[D_2]$					
		25	30	35	40	45	50
π	w/o inv.	80.25	93.04	102.61	111.61	120.61	129.61
	w/ inv.	90.66	104.16	117.66	129.91	139.97	148.97
q_1	w/o inv.	35.00	35.93	36.72	36.72	36.72	36.72
	w/ inv.	45.41	45.41	45.55	48.99	52.49	52.53
q_2	w/o inv.	0.00	0.00	2.33	7.33	12.33	17.33
	w/ inv.	0.00	0.00	0.00	0.00	0.00	4.95
$\hat{q}_2$	w/o inv.	15.00	17.37	17.67	17.67	17.67	17.67
	w/ inv.	10.83	15.84	20.66	21.62	22.48	22.48
I_1	w/o inv.	–	–	–	–	–	–
	w/ inv.	4.17	4.16	4.22	5.75	7.56	7.58

		Expected second period demand $E[D_2]$				
		55	60	65	70	75
π	w/o inv.	138.61	147.61	156.61	165.61	174.61
	w/ inv.	157.97	166.97	175.97	184.97	193.97
q_1	w/o inv.	36.72	36.72	36.72	36.72	36.72
	w/ inv.	52.53	52.53	52.53	52.53	52.53
q_2	w/o inv.	22.33	27.33	32.33	37.33	42.33
	w/ inv.	9.94	14.95	19.95	24.95	29.95
$\hat{q}_2$	w/o inv.	17.67	17.67	17.67	17.67	17.67
	w/ inv.	22.48	22.48	22.48	22.48	22.48
I_1	w/o inv.	–	–	–	–	–
	w/ inv.	7.58	7.58	7.58	7.58	7.58

9.5 Conclusion

In this paper, we have presented a stochastic single-product two-period newsvendor model to analyze the optimal production and remanufacturing decisions of a firm under uncertain demand. Our model extends previous research by jointly considering demand uncertainty, the possibility of keeping inventories of first period production and the fact that the supply of returned cores depends on manufacturing and supply decisions for brand new products in the past.

We analytically derive conditions on the optimality of strategies like excessive first period production, new production in period two, and remanufacturing. Using a numerical analysis we present sensitivity analysis, of the results with respect to model parameters like core collection yield, remanufacturing cost savings, and second period expected demand.

One of our main results is that contrary to some findings in the existing literature remanufacturing may not take place when costs for storage are relatively low.

Further we find that depending on the particular characteristics of the scenario studied, inventories and remanufacturing may either be substitutes or complements.

The former effect is observed when the core collection yield γ is large, while the latter effect occurs when core collection yield γ is small.

Finally, a third interesting observation is that excess first period production first increases and then decreases when the core collection yield γ increases. This is due to two opposite effects. First an increase in γ lowers the level of excess first period production necessary to obtain the same level of returned cores. Second, an increase in returned cores enables increased remanufacturing to substitute the more costly new production in period 2. Obviously, the latter effect explains the increase in excess new production in period 1 for small levels of γ when new period 2 production is necessary to achieve a sufficient second period supply. The former effect takes over, when the level of remanufacturing is already sufficient to substitute all new production in period 2.

Our work can be extended in several interesting directions. First, the core acquisition process seems to be an interesting area to study in detail. Following some models in the literature the amount of returned cores can be linked to an acquisition price function.

Second, in the same context it should be promising to include the quality of returned cores to allow for a more comprehensive analysis of the remanufacturing profitability. Third, following some of the literature, an extension toward a setting where manufacturers compete for the market of new and/or remanufactured products is interesting to focus on the strategic decisions whether and how to set up the closed-loop supply chains under different market environments.

Appendix

Proof of Proposition 1. Let λ^{R} correspond to the lagrangian multiplier of the remanufacturing constraint, while λ^{q_2} and $\lambda^{\hat{q}_2}$ correspond to the lagrangian multipliers of the nonnegativity constraints for q_2 and $\hat{q}_2$, respectively. Note that we do not need the nonnegativity constraint for first period production q_1 since this will be trivially $q_1 \geq 0$ due to our assumption $p_1 > c_1$ and the nonnegativity of demand. Then the system of Karush-Kuhn-Tucker (KKT) conditions is given by

$$-c_1 + (p_1 + \lambda^{\mathrm{R}} \gamma)\,[1 - F_{D_1}(q_1)] = 0, \tag{9.16}$$

$$-\beta\, c_2 + \beta\, p_2\,[1 - F_{D_2}(q_2 + \hat{q}_2)] + \lambda^{q_2} = 0, \tag{9.17}$$

$$-\beta\, c_2 + \beta\,\delta + \beta\, p_2\,[1 - F_{D_2}(q_2 + \hat{q}_2)] - \lambda^{\mathrm{R}} + \lambda^{\hat{q}_2} = 0, \tag{9.18}$$

$$\lambda^{\mathrm{R}}\,[\hat{q}_2 - \gamma S_{D_1}(q_1)] = 0, \tag{9.19}$$

$$\hat{q}_2 \leq \gamma S_{D_1}(q_1), \tag{9.20}$$

$$q_2\,\lambda^{q_2} = 0, \tag{9.21}$$

$$-q_2 \leq 0, \tag{9.22}$$

$$\hat{q}_2\,\lambda^{\hat{q}_2} = 0, \tag{9.23}$$

$$-\hat{q}_2 \leq 0, \tag{9.24}$$

$$\lambda^C \geq 0, \tag{9.25}$$

$$\lambda^{q_2}, \lambda^{\hat{q}_2} \geq 0. \tag{9.26}$$

Case 1. Exclusive, but limited remanufacturing in period 2:

This case implies that $\lambda^{\hat{q}_2} = 0$ and $\lambda^{q_2} > 0$. Further, since $\hat{q}_2 < \gamma S_{D_1}(q_1)$ it follows from (9.19) that $\lambda^{\mathrm{R}} = 0$.

Consequently, (9.17) leads to

$$F_{D_2}(\hat{q}_2) = \frac{\beta\,(p_2 - c_2) + \lambda^{q_2}}{\beta\, p_2}, \tag{9.27}$$

while (9.18) leads to

$$F_{D_2}(\hat{q}_2) = \frac{\beta\,(p_2 - c_2 + \delta)}{\beta\, p_2} = \frac{p_2 - c_2 + \delta}{p_2}. \tag{9.28}$$

From (9.27) and (9.28), we obtain $\lambda^{q_2} = \beta\,\delta > 0$. Finally, (9.16) leads directly to

$$F_{D_1}(q_1) = \frac{p_1 - c_1}{p_1}. \tag{9.29}$$

This completes the proof of case 1.

Case 2. Exclusive, full remanufacturing in period 2:

In this case, $\lambda^C > 0$ whereas $\lambda^{\hat{q}_2} = 0$ and $\lambda^{q_2} > 0$. While (9.17) again gives rise to (9.27), (9.18) now yields

$$F_{D_2}(\hat{q}_2) = \frac{\beta\,(p_2 - c_2 + \delta) - \lambda^{\mathrm{R}}}{\beta\, p_2}, \tag{9.30}$$

which—together with (9.27)—gives

$$\lambda^{\mathrm{R}} = \beta\,\delta - \lambda^{q_2}. \tag{9.31}$$

Since $\lambda^{q_2} > 0$, the upper boundary $\lambda^{\mathrm{R}} < \beta\,\delta$ follows directly. Finally, (9.16) leads directly to

$$F_{D_1}(q_1) = \frac{p_1 - c_1 + \lambda^{\mathrm{R}}\gamma}{p_1 + \lambda^{\mathrm{R}}\gamma}. \tag{9.32}$$

This completes the proof of case 2.

Case 3. Full remanufacturing and new production in period 2:

This case is induced by $\lambda^{\mathrm{R}} > 0$ as well as $\lambda^{q_2} = \lambda^{\hat{q}_2} = 0$. Then, (9.17) yields

$$F_{D_2}(q_2+\hat{q}_2) = \frac{\beta\,(p_2-c_2)}{\beta\,p_2} = \frac{p_2-c_2}{p_2}, \tag{9.33}$$

while (9.18) once again gives rise to

$$F_{D_2}(q_2+\hat{q}_2) = \frac{\beta\,(p_2-c_2+\delta)-\lambda^{\mathrm{R}}}{\beta\,p_2} \tag{9.34}$$

and (9.16) leads to (9.32). From (9.33) and (9.34), it follows that $\lambda^{\mathrm{R}} = \beta\,\delta$ and as a result (9.32) can be rewritten as

$$F_{D_1}(q_1) = \frac{p_1-c_1+\gamma\beta\,\delta}{p_1+\gamma\beta\,\delta}, \tag{9.35}$$

which concludes the proof of this third case.

□

Proof of Lemma 1. The proof of (9.4) follows directly from Proposition 1 and its proof. Consider the case $\lambda^{\mathrm{R}} = 0$. In that case, we observe from (9.29) that the optimal first period decision corresponds to the well-known single-period newsvendor quantity, denoted by q_1^{NV}. Moreover, the optimal second period supply comes exclusively from remanufacturing and is given by (9.27). We will denote this quantity by $\hat{q}_2^{\max}$. Since $\lambda^{\mathrm{R}} = 0$, this second period supply is unconstrained, which implies that $\hat{q}_2^{\max} \leq \gamma S_{D_1}(q_1^{\mathrm{NV}})$. On the other hand, whenever $\lambda^{\mathrm{R}} > 0$ we know from (9.19) that $\hat{q}_2 = \gamma S_{D_1}(q_1)$. Further, from (9.32) we observe that $q_1 > q_1^{\mathrm{NV}}$ which implies that $S_{D_1}(q_1) > S_{D_1}(q_1^{\mathrm{NV}})$, while (9.30) yields $\hat{q}_2 < \hat{q}_2^{\max}$. These conditions jointly hold only if $\hat{q}_2^{\max} > \gamma S_{D_1}(q_1^{\mathrm{NV}})$ which concludes the proof of this part of Lemma 1.

The proof of (9.6) follows directly from the proof of case 3 in Proposition 1. To prove (9.5) and (9.7), we need to consider the explicit formulae of $S_{D_1}(q_1)$ and $F_{D_t}^{-1}(q_t)$ associated with the uniform demand distributions. The expected sales are given by $S_{D_1}(q_1) = q_1 - \frac{(q_1-a_1)^2}{2\,(b_1-a_1)}$. The supply quantity is given by $q_t = F_{D_t}^{-1}(q_t) = a_t + (b_t-a_t)\,F_{D_t}(q_t)$. The rest is achieved by simple algebra. □

Proof of Proposition 2. Let λ^{R} and λ^{I} correspond to the lagrangian multipliers of the remanufacturing and inventory constraint, respectively. Further let λ^{I1}, λ^{q_2} and $\lambda^{\hat{q}_2}$ correspond to the lagrangian multipliers of the non-negativity constraints for I_1, q_2 and $\hat{q}_2$, respectively. Note that we do not need the nonnegativity constraint for first period production q_1 since this will be trivially $q_1 \geq 0$ due to our assumption $p_1 > c_1$ and the nonnegativity of demand. Then the system of KKT conditions is given by

$$-c_1 + (p_1+\lambda^{\mathrm{R}}\,\gamma-\lambda^{\mathrm{I}})\,[1-F_{D_1}(q_1)] + \lambda^{\mathrm{I}} = 0, \tag{9.36}$$

$$-h + \beta\,p_2\,[1-F_{D_2}(q_2+\hat{q}_2+I_1)] - \lambda^{\mathrm{I}} + \lambda^{I_1} = 0, \tag{9.37}$$

$$-\beta\, c_2 + \beta\, p_2\, [1 - F_{D_2}(q_2 + \hat{q}_2 + I_1)] + \lambda^{q_2} = 0, \tag{9.38}$$

$$-\beta\, c_2 + \beta\, \delta + \beta\, p_2\, [1 - F_{D_2}(q_2 + \hat{q}_2 + I_1)] - \lambda^{\mathrm{R}} + \lambda^{\hat{q}_2} = 0, \tag{9.39}$$

$$\lambda^{\mathrm{R}}\, [\hat{q}_2 - \gamma S_{D_1}(q_1)] = 0, \tag{9.40}$$

$$\hat{q}_2 \leq \gamma S_{D_1}(q_1), \tag{9.41}$$

$$\lambda^{\mathrm{I}}\, [I_1 - q_1 + S_{D_1}(q_1)] = 0, \tag{9.42}$$

$$I_1 \leq q_1 - S_{D_1}(q_1), \tag{9.43}$$

$$q_2\, \lambda^{q_2} = 0, \tag{9.44}$$

$$-q_2 \leq 0, \tag{9.45}$$

$$\hat{q}_2\, \lambda^{\hat{q}_2} = 0, \tag{9.46}$$

$$-\hat{q}_2 \leq 0, \tag{9.47}$$

$$I_1\, \lambda^{I_1} = 0, \tag{9.48}$$

$$-I_1 \leq 0, \tag{9.49}$$

$$\lambda^{C} \geq 0, \tag{9.50}$$

$$\lambda^{q_2}, \lambda^{\hat{q}_2}, \lambda^{I_1} \geq 0. \tag{9.51}$$

Case 1. Exclusive, but limited remanufacturing in period 2:

This case implies that $\lambda^{\hat{q}_2} = 0$, $\lambda^{q_2} > 0$, and $\lambda^{I_1} > 0$. Further, since $\hat{q}_2 < \gamma S_{D_1}(q_1)$ it follows from (9.41) that $\lambda^{\mathrm{R}} = 0$. Finally, $\lambda^{\mathrm{I}} = 0$ since $I_1 = 0 \leq q_1 - S_{D_1}(q_1)$.

Consequently, (9.36), (9.38), and (9.39) lead again to (9.27), (9.28), and (9.29) while (9.37) gives rise to

$$F_{D_2}(\hat{q}_2) = \frac{\beta\, p_2 - h + \lambda^{I_1}}{\beta\, p_2}. \tag{9.52}$$

From (9.28) and (9.52), it follows that $h = \beta\, (c_2 - \delta) + \lambda^{I_1}$. Since $\lambda^{I_1} > 0$, this implies that $h > \beta\, (c_2 - \delta)$ which concludes the proof of this part.

Case 2. Exclusive, full remanufacturing in period 2:

In this case, $\lambda^{C} > 0$ whereas $\lambda^{\mathrm{I}} = 0$, $\lambda^{\hat{q}_2} = 0$, $\lambda^{q_2} > 0$ and $\lambda^{I_1} > 0$. Once again the first part of the proof is identical to the proof of case 2 in Proposition 1. Further, (9.37) again gives rise to (9.52). Thus, from (9.30) and (9.52) we get

$$\lambda^{\mathrm{R}} = h - \beta\, (c_2 - \delta) - \lambda^{I_1} > 0, \tag{9.53}$$

which implies $h > \beta\, (c_2 - \delta)$ and completes the proof of case 2.

Case 3. Full remanufacturing and limited use of inventory in period 2:

This case is induced by $\lambda^{\mathrm{R}} > 0$, $\lambda^{q_2} > 0$ as well as $\lambda^{\mathrm{I}} = \lambda^{\hat{q}_2} = \lambda^{I_1} = 0$. In this case, (9.36) gives rise to (9.32), while (9.38) leads to

$$F_{D_2}(\hat{q}_2 + I_1) = \frac{\beta\,(p_2 - c_2) + \lambda^{q_2}}{\beta p_2}. \tag{9.54}$$

From (9.39), we obtain

$$F_{D_2}(\hat{q}_2 + I_1) = \frac{\beta\,(p_2 - c_2 + \delta) - \lambda^{\mathrm{R}}}{\beta p_2} \tag{9.55}$$

and (9.37) gives rise to

$$F_{D_2}(\hat{q}_2 + I_1) = \frac{\beta\, p_2 - h}{\beta\, p_2}. \tag{9.56}$$

Together, (9.54)–(9.56) yield $\lambda^{\mathrm{R}} = h - \beta\,(c_2 - \delta)$ and $\lambda^{\mathrm{R}} > 0$ implies that $h > \beta\,(c_2 - \delta)$ which concludes the proof of this part.

Case 4. Full remanufacturing and full use of inventory in period 2:

In this case $\lambda^{\mathrm{R}} > 0$, $\lambda^{\mathrm{I}} > 0$ and $\lambda^{q_2} > 0$ while $\lambda^{\hat{q}_2} = \lambda^{I_1} = 0$. Consequently, from (9.36) we obtain

$$F_{D_1}(q_1) = \frac{p_1 - c_1 + \lambda^{\mathrm{R}}\gamma}{p_1 + \lambda^{\mathrm{R}}\gamma - \lambda^{\mathrm{I}}}. \tag{9.57}$$

From (9.38) and (9.39) we once again obtain (9.54) and (9.55), respectively. Equation (9.37) now gives

$$F_{D_2}(\hat{q}_2 + I_1) = \frac{\beta\, p_2 - h - \lambda^{\mathrm{I}}}{\beta\, p_2}. \tag{9.58}$$

From (9.54) and (9.55), we get $\lambda^{\mathrm{R}} = \beta\,\delta - \lambda^{q_2}$ which—due to $\lambda^{q_2} > 0$—implies $\lambda^{\mathrm{R}} < \beta\,\delta$. Moreover, from (9.55) and (9.58) we obtain $\lambda^{\mathrm{R}} = h - \beta\,(c_2 - \delta) + \lambda^{\mathrm{I}}$. Since $\lambda^{\mathrm{I}} > 0$, it follows that $\lambda^{\mathrm{R}} > h - \beta\,(c_2 - \delta)$, which concludes the proof of this case.

Case 5. Full remanufacturing, full use of inventory and new production in period 2.

This case is characterized by $\lambda^{\mathrm{R}} > 0$ and $\lambda^{\mathrm{I}} > 0$, while $\lambda^{q_2} = \lambda^{\hat{q}_2} = \lambda^{I_1} = 0$. As in case 4 above, (9.36) gives rise to (9.57). From (9.37), we get

$$F_{D_2}(q_2 + \hat{q}_2 + I_1) = \frac{\beta\, p_2 - h - \lambda^{\mathrm{I}}}{\beta\, p_2}, \tag{9.59}$$

(9.38) yields

$$F_{D_2}(q_2+\hat{q}_2+I_1)=\frac{\beta\,(p_2-c_2)}{\beta\,p_2}=\frac{p_2-c_2}{p_2},\tag{9.60}$$

while (9.38) gives

$$F_{D_2}(q_2+\hat{q}_2+I_1)=\frac{\beta\,(p_2-c_2+\delta)-\lambda^{\mathrm{R}}}{\beta\,p_2}.\tag{9.61}$$

From (9.60) and (9.61), we get $\lambda^{\mathrm{R}}=\beta\,\delta$. Further from (9.59) and (9.60), we obtain $\lambda^{\mathrm{I}}=\beta\,c_2-h$ which implies $h<\beta\,c_2$ since $\lambda^{\mathrm{I}}>0$. Finally, note that using the expressions for λ^{I} and λ^{R} and our assumption $c_1+h>\beta\,c_2$ we get $0\le F_{D_1}(q_1)\le 1$ which concludes the proof of case 5.

Case 6. Exclusive, but limited use of inventory in period 2:

This case implies that $\lambda^{\hat{q}_2}>0$, $\lambda^{q_2}>0$ and $\lambda^{I_1}=0$. Further, since $0=\hat{q}_2<\gamma S_{D_1}(q_1)$ it follows from (9.19) that $\lambda^{\mathrm{R}}=0$. Finally, $\lambda^{\mathrm{I}}=0$ since $I_1\le q_1-S_{D_1}(q_1)$.

Consequently, (9.36) leads again to (9.29). Further, (9.37) gives rise to

$$F_{D_2}(I_1)=\frac{\beta\,p_2-h}{\beta\,p_2},\tag{9.62}$$

(9.38) yields

$$F_{D_2}(I_1)=\frac{\beta\,(p_2-c_2)+\lambda^{q_2}}{\beta\,p_2},\tag{9.63}$$

and from (9.39), we obtain

$$F_{D_2}(I_1)=\frac{\beta\,(p_2-c_2+\delta)+\lambda^{\hat{q}_2}}{\beta\,p_2}.\tag{9.64}$$

Finally, together (9.62) and (9.64) imply that $h=\beta\,(c_2-\delta)-\lambda^{\hat{q}_2}$. Since $\lambda^{\hat{q}_2}>0$, this implies that $h\le\beta\,(c_2-\delta)$ which concludes the proof of this part.

Case 7. Exclusive, full use of inventory in period 2:

In this case, $\lambda^{\mathrm{R}}=\lambda^{I_1}=0$ while $\lambda^{\mathrm{I}}>0$, $\lambda^{q_2}>0$, and $\lambda^{\hat{q}_2}>0$. Thus, (9.36) gives

$$F_{D_1}(q_1)=\frac{p_1-c_1}{p_1-\lambda^{\mathrm{I}}},\tag{9.65}$$

while (9.37) leads to

$$F_{D_2}(I_1)=\frac{\beta\,p_2-h-\lambda^{\mathrm{I}}}{\beta\,p_2}.\tag{9.66}$$

From (9.38) and (9.39), we obtain once again (9.63) and (9.64), respectively. As a result, it follows from (9.64) and (9.66) that $\lambda^{\mathrm{I}} = \beta\,(c_2 - \delta) - h - \lambda^{\hat{q}_2}$. Since $\lambda^{\hat{q}_2} > 0$, this gives the upper bound of λ^{I}. Further, by rewriting this term we get $h = \beta\,(c_2 - \delta) - \lambda^{\mathrm{I}} - \lambda^{\hat{q}_2}$. Clearly, this implies that $h \leq \beta\,(c_2 - \delta)$ which concludes the proof of this case.

Case 8. Full use of inventory and limited remanufacturing in period 2:

In this case, $\lambda^{\mathrm{R}} = \lambda^{I_1} = \lambda^{\hat{q}_2} = 0$ while $\lambda^{\mathrm{I}} > 0$ and $\lambda^{q_2} > 0$. From (9.36), we once again obtain (9.65) while (9.37) leads to

$$F_{D_2}(I_1 + \hat{q}_2) = \frac{\beta\, p_2 - h - \lambda^{\mathrm{I}}}{\beta\, p_2}, \tag{9.67}$$

(9.38) yields

$$F_{D_2}(I_1 + \hat{q}_2) = \frac{\beta\,(p_2 - c_2) + \lambda^{q_2}}{\beta\, p_2}, \tag{9.68}$$

and from (9.39) we obtain

$$F_{D_2}(I_1 + \hat{q}_2) = \frac{\beta\,(p_2 - c_2 + \delta)}{\beta\, p_2}. \tag{9.69}$$

From (9.67) and (9.69), we get $\lambda^{\mathrm{I}} = \beta\,(c_2 - \delta) - h$. Since $\lambda^{\mathrm{I}} > 0$, this implies that $h \leq \beta\,(c_2 - \delta)$ which concludes the proof of this case.

Case 9. Full use of inventory and full remanufacturing in period 2:

This case is identical to case 4. From (9.54) and (9.58), we get $\lambda^{\mathrm{I}} = \beta\, c_2 - h - \lambda^{q_2}$, which given that $\lambda^{q_2} > 0$ yields the upper bound on λ^{I}. Further, from (9.55) and (9.58) we obtain $\lambda^{\mathrm{I}} = \beta\,(c_2 - \delta) - h + \lambda^{\mathrm{R}}$ which—given that $\lambda^{\mathrm{R}} > 0$—directly yields the lower bound on λ^{I} and concludes the proof of this case.

Case 10. Full use of inventory, full remanufacturing, and new production in period 2:

This case and its proof is identical to case 5.

□

Proof of Lemma 2. For $h > h^*$, the proof of (9.12) is identical to the proof of the first part of Lemma 1. For $h \leq h^*$, consider the case $\lambda^{\mathrm{I}} = 0$. In that case, we observe from (9.29) that the optimal first period decision corresponds to the well-known single-period newsvendor quantity, denoted by q_1^{NV}. Moreover, the optimal second period supply comes exclusively from inventory and is given by (9.62). We will denote this quantity by $I_1^{\max}$. Since $\lambda^{\mathrm{I}} = 0$, this second period supply is unconstrained, which implies that $I_1^{\max} \leq q_1 - S_{D_1}(q_1^{\mathrm{NV}})$. On the other hand, whenever $\lambda^{\mathrm{I}} > 0$ we know from (9.42) that $I_1 = q_1 - S_{D_1}(q_1)$. Further, from

(9.65) we observe that $q_1 > q_1^{NV}$ which implies that $S_{D_1}(q_1) > S_{D_1}(q_1^{NV})$. Since at most the complete additional production can be sold we get $q_1 - S_{D_1}(q_1) \geq q_1^{NV} - S_{D_1}(q_1^{NV})$. Further, (9.66) yields $I_1 < I_1^{max}$. These conditions jointly hold only if $I_1^{max} > q_1 - S_{D_1}(q_1^{NV}) = I_{D_1}(q_1^{NV})$ which concludes the proof of this part of Lemma 2.

Let us now turn to the proof of condition (9.14). From the proof of Proposition 2, we know that no remanufacturing can only occur if $h > h^*$. Whenever $h \leq h^*$, we observe from (9.28) and (9.62) that $I_1^{max} \geq \hat{q}_2^{max}$. From case 7 of Proposition 2, we know that whenever inventory is fully used $I_1 = F_{D_2}^{-1}(\frac{\beta p_2 - h - \lambda^I}{\beta p_2}) = I_{D_1}(q_1) \leq I_1^{max}$. From case 8 of Proposition 2, we observe that remanufacturing occurs as soon as $\lambda^I \geq \beta(c_2 - \delta) - h$ which implies that the associated inventory level $I_{D_1}(q_1) \leq \hat{q}_2^{max}$. Thus no remanufacturing occurs whenever $I_{D_1}(q_1) > \hat{q}_2^{max}$.

The proof of (9.15) follows directly from the proof of case 5 of Proposition 2.

□

References

Bhattacharya, S., Guide, V. D. R., & Van Wassenhove, L. N. (2006). Optimal order quantities with remanufacturing across new product generations. *Production and Operations Management*, *15*, 421–431.

Cattani, K.D., Dahan, E., & Schmidt, G.M. (2008). Tailored capacity: speculative and reactive fabrication of fashion goods. *International Journal of Production Economics*, *114*, 416–430.

Chung, C.-S., Flynn, J., & Kirca, O. (2008). A multi-item newsvendor problem with preseason production and capacitated reactive production. *European Journal of Operational Research*, *188*, 775–792.

Ferguson, M., & Toktay, L. B. (2006). The effect of competition on recovery strategies. *Production and Operations Management*, *15*, 351–368.

Ferguson, M., Guide Jr., V.D.R., Koca, E., & Souza, G.C. (2009). The value of quality grading in remanufacturing. *Production and Operations Management*, *18*, 300–314.

Ferrer, G., & Swaminathan, J.M. (2006). Managing new and remanufactured products. *Management Science*, *52*, 15–26.

Ferrer, G., & Swaminathan, J.M. (2010). Managing new and differentiated remanufactured products. *European Journal of Operational Research*, *203*, 370–379.

Galbreth, M., & Blackburn, J. (2006). Optimal acquisition and sorting policies for remanufacturing. *Production and Operations Management*, *15*, 384–392.

Granot, D., & Yin, S. (2008). Price and order postponement in a decentralized newsvendor model with multiplicative and price-dependent demand. *Operations Research*, *56*, 121–139.

Guide Jr., V.D.R., & Li, J. (2010). The potential for cannibalization of new product sales by remanufactured products. *Decision Sciences*, *41*, 547–572.

Guide Jr., V.D.R., & Van Wassenhove, L.N. (2009). The evolution of closed-loop supply chain research. *Operations Research*, *57*, 10–18.

Guide Jr., V.D.R., Souza, G.C., Van Wassenhove, L.N., & Blackburn, J.D. (2006). Time value of commercial product returns. *Management Science*, *52*, 1200–1214.

Inderfurth, K. (2004). Optimal policies in hybrid manufacturing/remanufacturing systems with product substitution. *International Journal of Production Economics*, *90*, 325–343.

Kelle, P., & Silver, E. A. (1989). Purchasing policy of new containers considering the random returns of previously issued containers. *IIE Transactions*, *21*, 349–354.

Li, X., Li, Y., & Cai, X. (2011). Quantity decisions in a supply chain with early returns remanufacturing. *International Journal of Production Research*, doi:10.1080/00207543.2011.565085.

Li, Y., Zhang, J., Chen, J., & Cai, X. (2010). Optimal solution structure for multi-period production planning with returned products remanufacturing. *Asia-Pacific Journal of Operational Research*, *5*, 629–648.

Olugu, E.U., Wong, K.Y., & Shaharoun, A.M. (2011). Development of key performance measures for the automobile green supply chain. *Resources, Conservation and Recycling*, *55*, 567–579.

Reimann, M. (2011a). Speculative production and anticipative reservation of reactive capacity by a multi-product newsvendor. *European Journal of Operational Research*, *211*, 35–46.

Reimann, M. (2011b). *Accurate response by postponement*, Impuls Working paper 01/2011, University of Graz.

Shi, J., Zhang, G., & Sha, J. (2011). Optimal production planning for a multi-product closed-loop system with uncertain demand and return. *Computers and Operations Research*, *38*, 641–650.

Teunter, R. H., & Flapper, S. D. P. (2011). Optimal core acquisition and remanufacturing policies under uncertain core quality fractions. *European Journal of Operational Research*, *210*, 241–248.

Wei, C., Li, Y., & Cai, X. (2011). Robust optimal policies of production and inventory with uncertain returns and demand. *International Journal of Production Economics*, *134*, 357–367.

Zhang, B., & Du, S. (2009). Multi-product newsboy problem with limited capacity and outsourcing. *European Journal of Operational Research*, *202*, 107–113.

Zhou, S., Tao, Z., & Chao, X. (2011). Optimal control of inventory systems with multiple types of remanufacturable products. *Manufacturing & Service Operations Management*, *13*, 20–34.

Zhou, S. X., & Yu, Y. (2011). Optimal product acquisition, pricing, and inventory management for systems with remanufacturing. *Operations Research*, *59*, 514–521.

Chapter 10
The Remanufacturing Newsvendor Problem

Matthew J. Drake, David W. Pentico, and Robert P. Sroufe

Abstract As companies have synchronized and streamlined their traditional supply chains to improve customer service while reducing costs, many firms have begun to consider the impact of end-of-life unit flows on their overall profitability. Many companies have begun to accept end-of-life items from their customers for corporate social responsibility or customer service reasons, and they must develop processes to handle this large volume of units. Some firms have chosen to capture value from these flows by remanufacturing them and selling the units in a secondary market. The quality of the end-of-life items acquired from customers, however, is quite variable, and some items may not be in good enough condition to remanufacture. Thus, the items must be inspected either one by one (discrete) or in a batch to assess their condition. We have developed newsvendor models to assist firms in making the inspection decision to yield maximum expected profit when faced with a fixed number of end-of-life items. Both inspection processes generate the same target inventory level to sell in the secondary market. We examined the difference between the batch and discrete inspection processes using simulation and found that the batch process yields a higher average profit because it can allow the firm to take advantage of large demands with higher than expected inspection yields.

Keywords Remanufacturing • Inspection decision • End-of-life items • Simulation • Batch process • Inspection yields

M.J. Drake (✉) • D.W. Pentico • R.P. Sroufe
Palumbo-Donahue Schools of Business, Duquesne University, Pittsburgh, PA 15282, USA
e-mail: drake987@duq.edu; pentico@duq.edu; sroufer@duq.edu

T.-M. Choi (ed.), *Handbook of Newsvendor Problems*, International Series in Operations Research & Management Science 176, DOI 10.1007/978-1-4614-3600-3_10,

10.1 Introduction

Traditional supply chain management initiatives tend to focus on improving the "forward" supply chain, which consists of the flow of goods and services downstream from suppliers to manufacturers, distributors, retailers, and ultimately end users. Over the past decade, however, many firms have begun to consider the reverse flows that inevitably occur from retailers and customers back to the manufacturers. These reverse flows are so significant that the Supply-Chain Council amended its existing Supply Chain Operations Reference (SCOR©) model processes of plan, source, make, and deliver in the early 2000s to include an additional basic supply chain process: return.

Reverse supply chain flows occur for a number of reasons. A major source of these flows is the products that customers return because of damages, shipping errors, buyback agreements, or customer service policies. Manufacturers have always had to deal with these returned items, and the volume of returned products can be quite significant depending on the industry. It is estimated that consumer returns in the United States represents approximately 6% of a firm's revenue, and the return rates can vary from 10% for many companies to over 40% for catalog and Internet retailers (Horvath et al. 2005; Rogers and Tibben-Lembke 2006). The costs associated with transporting, inspecting, and processing returns can be significant as well; for some products, the cost can be $150 or more per unit (Wyld 2004). Firms can even realize some hidden costs associated with returned items such as poor warehouse space utilization and higher taxes on inventory (Mannella 2003). Not surprisingly, many companies traditionally viewed their return operations as a cost center and designed their operations to process these items at the minimum possible cost.

In recent years, end-of-life items have become another prominent reverse flow in the supply chain for many manufacturers. In some instances, such as in Europe after the European Union adopted the Waste Electrical and Electronic Equipment (WEEE) legislation, companies are forced by governments to collect these end-of-life items and dispose of them in ways other than simply depositing them in a landfill (Mollenkopf 2006). In countries, such as the United States, that have not legislated manufacturer responsibility for end-of-life items, innovative companies are accepting or proactively acquiring end-of-life items with the goal of capturing value. This value could come through enhancement of the company's brand as being "sustainable" or "eco-friendly" or from more tangible processes such as remanufacturing or materials reclamation that increase the firm's profitability (Drake and Ferguson 2008).

Firms seeking to create value from their reverse supply chain flows must develop an efficient way to inspect and process their incoming items quickly and effectively. Depending on the quality of the items received, which can vary greatly for end-of-life items, the company has a set of viable disposal options. It is critical for firms to be able to direct specific units to the most appropriate disposal option that generates the maximum value possible given the unit's condition. Some of these options such as donations to charitable organizations, recycling, or reclaiming materials from the product have limited value potential; whereas, reconditioning or remanufacturing the product and selling it in a secondary market could bring significantly more value. A portion of the returned items will ultimately end up in a landfill if there are no other viable alternatives (Rogers and Tibben-Lembke 1999).

For items with particularly short life cycles such as many consumer electronics goods, the speed of quality assessment and remanufacturing can be crucial in maximizing the value that companies can capture in the secondary market for refurbished or remanufactured goods. Blackburn et al. (2004) estimate that the value of these short-life-cycle items can fall by more than 1% per week, and this rate can increase as the product reaches the end of its useful life. They note that Hewlett-Packard notebook computers spend approximately 18 weeks in transit and refurbishing before they reach the secondary market; thus, the computers have lost 20% or more of their potential value due to this delay (Guide et al. 2005). The time-sensitive value of short-life-cycle products also suggests that firms only have one reasonable opportunity to supply the secondary market with remanufactured units of a particular model or style. This factor led us to model inventory decisions in this environment with the newsvendor framework.

While some firms such as Bosch tools have proactively installed counters, sensors, or data loggers in their products to help their personnel assess the quality of an end-of-life item quickly and easily, the quality inspection process for reverse flows can be costly for many companies (Krikke 2001). The products' quality themselves can be highly variable, from like-new to unsalvageable. If the inspection process is costly and difficult, firms may want to limit the number of units that they inspect to avoid these large expenses as well as a result of the variability in the quantity of high-valued, remanufacturable units that the process yields.

In this chapter, we model the firm's inspection decision when faced with a fixed quantity of returned items. We determine the optimal inspection policy under batch and discrete assessment processes to maximize expected profit that the firm earns in the secondary market. The next section discusses our models in the context of relevant literature. Section 10.3 develops the two models, and Sect. 10.4 provides some simulation results on several numerical examples. The chapter concludes with some conclusions and discussion of future research possibilities.

10.2 Literature Review

The models we develop in this chapter are related to several distinct bodies of literature. The models are adaptations and extensions of the classic newsvendor models for single-period inventory control decisions. One such model is that by Mostard and Teunter (2006), which extends the traditional newsvendor model to include the possibility of receiving items returned by customers that are in good enough condition to resell in the primary market. Khouja (1999) provides a review of almost 100 newsvendor models and classifies the extensions to the classic model into 11 categories where one or more of the original assumptions have been relaxed or modified. Our model fits into several of these categories including models that consider multiple products and models with random yield.

There is a broad class of newsvendor models that generate single-period ordering or production quantities for multiple products subject to a set of constraints. These constraints often represent budget or capacity limitations (e.g., Abdel-Malek et al. 2004; Abdel-Malek and Montanari 2005; Abdel-Malek and Areeratchakul 2007; Ben-Daya and Raouf 1993; Erlebacher 2000; Lau and Lau 1995; Lau and Lau 1996; Moon and Silver 2000; Niederhoff 2007; Vairaktarakis 2000). Our models do include multiple products (remanufacturable units, nonremanufacturable units, and items that the company chooses not to inspect) with capacity constrained by the total end-of-life items available, but we consider random yield as well because the quality of the inspected units is variable.

A related model by Grubbström (2010) models the newsvendor problem using a compound demand process. Customers arrive according to a renewal process, and their quantity demanded when they arrive follows a second independent demand distribution. This model includes two demand distributions just as our model does, but our one source of uncertainty is remanufacturing yield rather than customer arrivals.

Another class of inventory models considers random production yield. Yano and Lee (1995) provide a review of many of these models. There have been a few studies that have treated random yield in a single-period newsvendor context such as those by Noori and Keller (1986), Inderfurth (2004), and Rekik et al. (2007). Yang et al. (2007) use the active set method and Newton's method to determine the optimal quantities to order from multiple suppliers each with its own random yield. These models, however, have not considered a capacitated environment as our model does. Ketzenberg et al. (2006) develop single- and multiple-period models for remanufacturing quantity decisions with random yield. The firm has the additional option of purchasing new units from a reliable supplier to meet demand, which is not an option in our model. They also assume that all returns are remanufactured, whereas our major decision variable is the number of returned items to inspect and consider for remanufacturing. The one model that does consider capacity limitations as well as random yield is that of Abdel-Malek et al. (2008). This model considers the problem faced by farmers or gardeners who have to determine the quantity of each product to plant in a given amount of farm land to minimize total cost (including lost revenue for demand that is not satisfied). Both the demand and the yield of each product is randomly distributed according to general probability functions. Our models differ from Abdel-Malek et al. (2008) in the decision and cost structures and in the fact that we consider both discrete and batch inspection decisions.

Our models are also related to a body of literature containing closed-loop supply chain inventory models. The models in these studies often simultaneously consider the tradeoffs between new production and remanufacturing operations to satisfy current demand. They also focus on the decisions related to the acquisition of end-of-life items that can be remanufactured through various incentive mechanisms.

See, for example, Savaskan et al. (2004), Aras et al. (2006), Bakal and Akcali (2006), Debo et al. (2006), Ferrer and Swaminathan (2006), Galbreth and Blackburn (2006), Mukhopadhyay and Ma (2009), and Ahiska and King (2010). These models treat a broader problem than ours does by looking at the interaction between new and remanufactured products in the market. Bhattacharya et al. (2006) develop a particularly relevant closed-loop newsvendor model for a retailer ordering new products from a manufacturer as well as remanufactured products from a prior product generation. They consider multi-period decisions and investigate channel coordination efforts among all three firms involved. They also assume that the new and remanufactured products are perfect substitutes in the consumer demand market. Our models apply to products that have a clear secondary market that is not a strong substitute for new items in the traditional market. We consider a separate demand for remanufactured products that is distinct from that of new products. Many of these models also assume that the remanufacturer has the ability to acquire as many end-of-life units as it wants; our model addresses a fixed supply of end-of-life items.

Inventory and production decisions for remanufacturing operations have become popular research topics in the past fifteen years. Many inventory models such as Teunter (2001), Mahadevan et al. (2003), Atasu and Cetinkaya (2006), and Tang and Teunter (2006) have been developed in recent years to accommodate remanufacturing operations, but these have focused on multiple-period, ongoing policies and not the single-period newsvendor-type decisions that are most appropriate for a short-life-cycle product market.

10.3 The Remanufacturing Newsvendor Models

Suppose that a firm has acquired Q end-of-life items. History has shown that ρ (where $\rho \in [0,1]$) proportion of the units are of high enough quality to be remanufactured effectively. The firm can inspect the quality of an end-of-life units at a cost of c_{i} per unit, and each unit is assumed to have a ρ percent chance of being remanufacturable. The inspection process in this model can accurately identify the quality of the end-of-life items, and the firm chooses to remanufacture all units of requisite quality. The unit cost of remanufacturing is c_{r}, and the remanufacturing process produces no defects or waste. The main decision variable in the model is k, the number of units to inspect (where $k \leq Q$).

Demand for remanufactured units in the secondary market is probabilistic with cumulative distribution function $F(x)$ and probability density function $f(x)$; remanufactured units sold in the secondary market yield a revenue of r per unit. We assume that the remanufacturing operation is a profitable one, meaning that the revenue obtained from selling a remanufactured unit is larger than the cost of inspecting and remanufacturing (i.e., $r > c_{\mathrm{i}} + c_{\mathrm{r}}$). Any remanufactured units that are not sold in the secondary market at the end of the selling season have a salvage value of s_1 per unit. End-of-life items that have not had their quality inspected

Table 10.1 Summary of model notation

Parameter	Description
r	Unit revenue for remanufactured item sold in secondary market
c_r	Unit remanufacturing cost
c_i	Unit inspection cost
s_1	Salvage value of unsold remanufactured unit
s_2	Salvage value of end-of-life item that was not inspected
s_3	Salvage value of nonremanufacturable end-of-life item
Q	Supply of end-of-life items available for inspection
k	Number of end-of-life items to be inspected (decision variable)

number $Q-k$ units and have a salvage value of s_2 per unit. Units that have been inspected and were found to be unsuitable for remanufacturing have a salvage value of s_3. As the value of these leftover units is dependent upon their utility and potential attractiveness in a salvage market, we assume that remanufactured units are the most valuable followed by units that have not been inspected (because they still hold the possibility that they could turn out to be remanufacturable); thus, $s_1 > s_2 > s_3$. This notation is summarized in Table 10.1.

When considering the decision of how many units to inspect, the firm has two general options available. It could decide on a fixed number of units to inspect *before* any inspections are performed and then remanufacture whichever units turn out to be of high enough quality. This would be appropriate if the inspection process is long or if all of the units can be inspected at once. We denote this situation as *batch inspection*. In other circumstances, however, it may be possible for the firm to inspect the end-of-life units one at a time and decide when to stop the inspection process based on the results of the previous inspections. We refer to this decision environment as *discrete inspection* because the firm decides whether to inspect each unit sequentially. We analyze each situation in the remainder of this section.

10.3.1 Batch Inspection Model

Under a batch inspection environment, the firm must determine the optimal number of end-of-life units to inspect before inspecting any of the units. As a result, we can analyze the inspection decision in a traditional newsvendor framework by developing the expected profit function and finding the value of $k \in [0, Q]$ where the first derivative equals zero. The firm's expected profit function for the batch inspection decision is:

$$\Pi(k) = r\left[\int_0^{\rho k} x f(x)\mathrm{d}x + \rho k(1-F(\rho k))\right] + s_1 \int_0^{\rho k} (\rho k - x) f(x)\mathrm{d}x$$
$$+ s_2(Q-k) + s_3(1-\rho)k - c_i k - c_r \rho k, \qquad (10.1)$$

where ρk represents the expected number of remanufacturable end-of-life units that are identified through the inspection process, $(1-\rho)k$ is the expected number of nonremanufacturable units, and $Q-k$ is the number of units that are not inspected. Taking the first derivative and setting it equal to zero yield the following critical ratio which represents the optimal batch inspection quantity:

$$k^* \in \left\{ k : F(\rho k) = \frac{\rho(r-c_r) + s_3(1-\rho) - s_2 - c_i}{\rho(r-s_1)} \right\}. \tag{10.2}$$

We can use the structure of the critical ratio to determine minimum and maximum values of ρ for which this quantity is optimal. The numerator of the fraction must be greater than or equal to zero for the equation to have a solution. Thus, we have

$$\rho^{\min} = \frac{s_2 + c_i - s_3}{r - c_r - s_3}. \tag{10.3}$$

For values of ρ less than $\rho^{\min}$, the optimal decision is for the firm to refrain from inspecting the quality of any of the end-of-life items. In this case, the likelihood of finding items of high enough quality to be remanufactured is too small in comparison with the given cost and revenue parameter values to support a remanufacturing operation.

The critical ratio must also be less than or equal to one in order for the equation in (10.2) to have a solution. Setting the numerator less than or equal to the denominator yields

$$\rho^{\max} = \frac{s_2 + c_i - s_3}{s_1 - c_r - s_3}. \tag{10.4}$$

For values of ρ greater than $\rho^{\max}$, the firm's optimal strategy is to inspect all Q of the end-of-life items that it has available. This is also the case if the value of k^* in (10.2) happens to be larger than Q.

10.3.2 Discrete Inspection Model

In the discrete inspection environment, the firm is able to inspect the quality of end-of-life units one by one; thus, the firm can stop the inspection process whenever it has the total number of units that it wants to remanufacture. In response to the structure of the decision environment, we analyze the optimal inspection decision using an iterative framework.

Suppose that you already have k remanufacturable units as a result of previous inspections and you want to decide whether or not to inspect one additional unit. The firm earns a marginal profit of s_2 from salvaging the unit with an unknown quality level. An inspection can yield two possible outcomes: a remanufacturable unit or a nonremanufacturable unit. If the inspection process finds that the unit is not remanufacturable, the firm earns a marginal profit of $s_3 - c_i$. If the unit is remanufacturable, on the other hand, that unit could either be demanded in the secondary market for revenue r, or it could not sell in the market and would only return salvage value s_1. Thus, the expected marginal profit if the unit is remanufacturable is

$$E[\Pi^{\text{Reman}}] = [r-(c_i+c_r)]P(D \geq k+1) + [s_1-(c_i+c_r)]P(D < k+1). \quad (10.5)$$

Combining these two possible expected marginal profit values, we can express the expected profit from inspecting the next end-of-life item as

$$E[\Pi^{\text{Inspect}}] = \{[r-(c_i+c_r)]P(D \geq k+1) + [s_1-(c_i+c_r)] \\ P(D < k+1)\}\rho + \{s_3 - c_i\}(1-\rho). \quad (10.6)$$

We can rewrite the expected marginal profit from inspecting the additional unit in (10.6) as

$$E[\Pi^{\text{Inspect}}] = \bar{r} - c_i - c_r\rho, \quad (10.7)$$

where $\bar{r} = [rP(D \geq k+1) + s_1 P(D < k+1)]\rho + s_3(1-\rho)$ represents the expected marginal revenue depending on the uncertain outcome of the inspection process and the demand for the remanufactured item in the secondary market. The firm should inspect the additional unit if

$$\bar{r} - c_i - c_r\rho \geq s_2. \quad (10.8)$$

Formally, the firm should stop the inspection process when either of the following two situations occur:

1. $\bar{r} < s_2 + c_i + c_r\rho$.
2. $k = Q$, which means that there are no more end-of-life items to inspect.

By rearranging the terms algebraically, we can write the first condition above as a critical-ratio-type expression:

$$P(D < k+1) = F(k) > \frac{\rho(r-c_r) + s_3(1-\rho) - s_2 - c_i}{\rho(r-s_1)}. \quad (10.9)$$

Note that the critical ratio for the discrete inspection process in (10.9) is identical to that of the batch inspection process developed in (10.2). The only element that differs between the two optimal inspection quantities is the argument of the cumulative distribution function of demand in the secondary market. The arguments both represent the number of remanufacturable units that the firm would like to have to sell in the secondary market. In the batch inspection process, however, the firm can only make its decision on the number of items to inspect (k) with the hope of having ρk units to remanufacture. In the discrete inspection environment, the firm knows exactly how many remanufacturable units that it currently has when it makes the decision about whether or not to inspect the next unit. These two optimal quantities may look different at first glance, but they are identical in that the firm wants to have the same quantity of remanufacturable units to sell in the secondary market.

10.4 Numerical Examples

The fact that the optimal inspection policy has the same target number of remanufacturable units in both batch and discrete inspection environments prompted us to investigate the behavior of the profit function in each scenario to see if the inspection structure has an impact on a firm's profit from remanufacturing operations. We built a simulation model in Crystal Ball (a spreadsheet add-in) to examine the firm's profit under batch and discrete inspection processes. This simulation model uses the same demand values and inspection results to compute profit realizations for batch and discrete inspections. The only part of each realization that differs is the inspection policy, which means that all observed differences in profit are attributed to the policy itself and not any other form of variation.

This section presents and discusses the results of two simulation studies performed with the base model. These two studies are not intended to provide a comprehensive analysis of the two inspection policies' performance overall; the goal is merely to illustrate that the two inspection processes generate different behavior in the profit function. A structured, comprehensive simulation analysis (to be completed in future research) is required to fully assess the impact of the inspection policy on the firm's expected profit. The two simulation studies are distinguished by their values of ρ to evaluate performance under high quality end-of-life items ($\rho = 0.75$) and low quality end-of-life items ($\rho = 0.35$). We ran 10,000 replications of each model to limit the standard error of our results. Demand in the secondary market was assumed to follow a normal distribution with mean 30 and standard deviation of 10, and the firm had $Q = 100$ end-of-life items that could be inspected. Table 10.2 provides the common revenue and cost parameter values used throughout the study.

Figures 10.1 and 10.2 display the results of the two simulation studies. In both scenarios, the batch inspection process yields a higher average profit than the discrete inspection process does, and this difference is statistically significant with a p-value less than 0.001. This may seem counterintuitive because discrete inspection allows the firm to react to the results of previous inspections whereas the batch inspection process requires that the firm make an irrevocable decision to inspect a fixed number of end-of-life items before seeing the results of any individual inspections.

The batch process has a higher average profit in both scenarios because it allows the firm to take advantage of favorable effects of randomness in some selling seasons. In the discrete environment, the firm will never have any more remanufacturable units to sell in the secondary market than the number prescribed

Table 10.2 Summary of simulation model parameter values

Parameter	Value	Parameter	Value
r	50	s_1	5
c_r	8	s_2	3
c_i	3	s_3	1

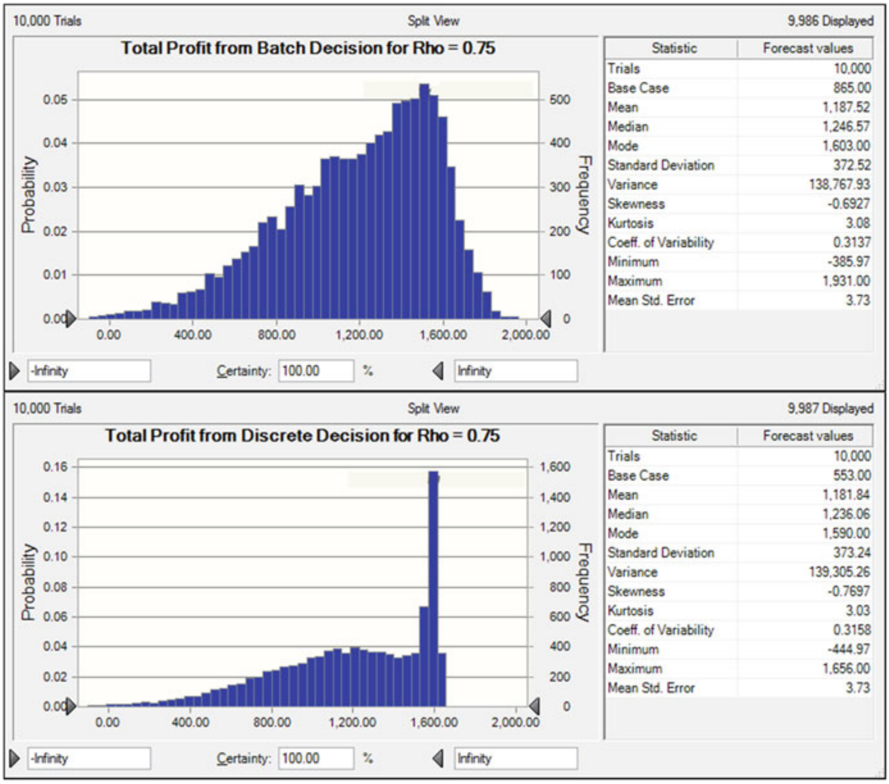


Fig. 10.1 Simulation results for batch and discrete inspection process with $\rho = 0.75$

by the critical ratio. The inspection process stops at this point (if the firm is maximizing expected profit). In the batch inspection environment, the firm tries to have the same number of remanufacturable units, but the actual number of units available to sell in the secondary market could easily be larger or fewer depending on the uncertain outcome of each inspection. If the firm lucks out and identifies a large number of remanufacturable units, and that happens to coincide with a large amount of demand in the secondary market, the firm will earn a large profit. The firm does not have the inventory to take advantage of large amounts of demand in the secondary market under the discrete inspection environment; thus, the upper bound of potential profit is smaller than in the batch case.

The same situation can occur at the other end of the demand spectrum. In periods where demand in the secondary market turns out to be low, it is possible for the firm in the batch environment to happen to identify only a small number of remanufacturable units. The firm would not have the large amount of leftover, remanufactured inventory at the end of the selling season that it would in the

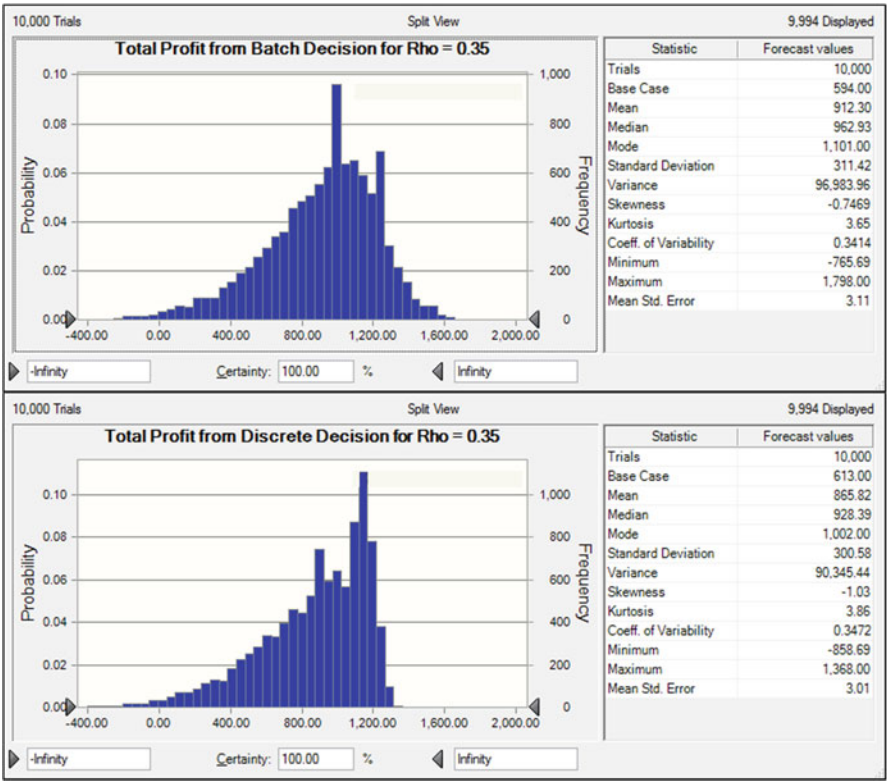


Fig. 10.2 Simulation results for batch and discrete inspection process with $\rho = 0.35$

discrete environment. The discrete inspection process almost always yields the critical ratio-specified number of remanufacturable units; the only way that the firm has a different number of remanufacturable units available is if it runs out of end-of-life items to inspect before it reaches the desired amount.

10.5 Conclusions and Future Research

In this chapter, we have developed a newsvendor inventory model to guide the inspection decision for remanufacturers with a fixed number of end-of-life items at hand. We analyzed the optimal inspection decision under both batch and discrete inspection environments, and we found that the target inventory level for the secondary market was identical for both inspection processes. Through a simulation study, we found that the batch inspection process yields a higher average profit compared with the discrete inspection process. This result is somewhat counterintuitive because the discrete process allows the firm to base the inspection decision for each individual unit on the history of previous inspections to that point; whereas, the batch inspection process requires that the firm decide to inspect a fixed number of units before observing any results. The batch process has a higher average profit because it gives the firm a chance to have a large number of remanufactured items to sell in the secondary market when demand happens to be high and vice versa; the discrete process, on the other hand, yields approximately the same number of remanufactured items in every selling season. A more complete simulation study under various problem parameter values (such as demand variance and the cost values) is required in the future to capture a full understanding of the relative performance of the two different inspection policies.

This study is the first in what could be a series of research studies related to inventory control for remanufactured items in the short-life-cycle market. A natural next step would be to relax the assumption in this study that the firm has a given number of end-of-life items available and consider an endogenous acquisition decision within the model. Depending on the proportion of high-quality items available as well as the acquisition cost, the firm may want to obtain more or fewer end-of-life items in the first place to maximize expected profit. It would also be interesting to investigate the efficacy of different pricing mechanisms to acquire the end-of-life items. If the remanufacturing firm was a subsidiary of the same company that originally manufactured the item, it may be possible to incorporate the remanufacturing operation into the initial production planning models for new items to create a comprehensive closed-loop supply chain model. One final extension to the base models in this chapter would be to introduce defects into the inspection or remanufacturing processes to analyze how the optimal inspection decision would change in response to increased process variability.

References

Abdel-Malek, L. L., & Areeratchakul, N. (2007). A quadratic programming approach to the multi-product newsvendor problem with side constraints. *European Journal of Operational Research*, *176*, 1607–1619.

Abdel-Malek, L. L., & Montanari, R. (2005). An analysis of the multi-product newsboy problem with a budget constraint. *International Journal of Production Economics*, *97*, 296–307.

Abdel-Malek, L. L., Montanari, R., & Meneghetti, D. (2008). The capacitated newsboy problem with random yield: the gardener problem. *International Journal of Production Economics*, *115*, 113–127.

Abdel-Malek, L. L., Montanari, R., & Morales, L. C. (2004). Exact, approximate, and generic iterative models for the Newsboy problem with a budget constraint. *International Journal of Production Economics*, *91*, 189–198.

Ahiska, S. S., & King, R. E. (2010). Inventory optimization in a one product recoverable manufacturing system. *International Journal of Production Economics*, *124*, 11–19.

Aras, N., Verter, V., & Boyaci, T. (2006). Coordination and priority decisions in hybrid manufacturing/remanufacturing systems. *Production and Operations Management*, *15*(4), 528–543.

Atasu, A., & Çetinkaya, S. (2006). Lot sizing for optimal collection and use of remanufacturable returns over a finite life-cycle. *Production and Operations Management*, *15*(4), 473–487.

Bakal, I. S., & Akcali, E. (2006). Effects of random yield in remanufacturing with price-sensitive supply and demand. *Production and Operations Management*, *15*(3), 407–420.

Ben-Daya, M., & Raouf, A. (1993). On the constrained multi-item single period inventory problem. *International Journal of Operations and Production Management*, *13*, 104–112.

Bhattacharya, S., Guide Jr., V. D. R., & Van Wassenhove, L. N. (2006). Optimal order quantities with remanufacturing across new product generations. *Production and Operations Management*, *15*(3), 421–431.

Blackburn, J.D., Guide Jr., V. D. R., Souza, G. C., & Van Wassenhove, L. N. (2004). Reverse supply chains for commercial returns. *California Management Review*, *46*(2):6–22.

Debo, L. G., L. B. Toktay, and L. N. Van Wassenhove (2006). Joint life-cycle dynamics of new and remanufactured products. *Production and Operations Management*, *15*(4):498–513.

Drake, M. J., & Ferguson, M. E. (2008). Closed-loop supply chain management for global sustainability. In C. Wankel & J. Stoner (Eds.) *Global sustainability initiatives: New models and new approaches* (pp. 173–191). Information Age Publishers: Charlotte.

Erlebacher, S. J. (2000). Optimal and heuristic solutions for the multi-item newsvendor problem with a single capacity constraint. *Production and Operations Management*, *9*(3), 303–318.

Ferrer, G., & Swaminathan, J. M. (2006). Managing new and remanufactured products. *Management Science*, *52*(1), 15–26.

Galbreth, M. R., & Blackburn, J. D. (2006). Optimal acquisition and sorting policies for remanufacturing. *Production and Operations Management*, *15*(3), 384–392.

Grubbström, R. W. (2010). The Newsboy problem when customer demand is a compound renewal process. *European Journal of Operational Research*, *203*, 134–142.

Guide Jr., V. D. R., Muyldermans, L., & Van Wassenhove, L. N. (2005). Hewlett-Packard company unlocks the value potential for time-sensitive returns. *Interfaces*, *35*(4), 281–293.

Horvath, P. A., Autry, C. W., & Wilcox, W. E. (2005). Liquidity implications of reverse logistics for retailers: a Markov chain approach. *Journal of Retailing*, *81*(3), 191–203.

Inderfurth, K. (2004). Analytical solution for a single-period production-inventory problem with uniformly distributed yield and demand. *Central European Journal of Operational Research*, *12*, 117–127.

Ketzenberg, M., van der Laan, E., and Teunter, R. H. (2006). Value of information in closed loop supply chains. *Production and Operations Management*, *15*(3), 393–406.

Khouja, M. (1999). The single-period (news-vendor) problem: literature review and suggestions for future research. *Omega*, *27*, 537–553.

Krikke, H. (2001). Recovery strategies and reverse logistics network design. In J. Sarkis, (Ed.). *Greener Manufacturing and Operations* (pp. 256–272). Greenleaf: Sheffield.

Lau, H. S., & Lau, A. H. L. (1995). The multi-product multi-constraint newsboy problem: application formulation and solution. *Journal of Operations Management*, *13*, 153–162.

Lau, H. S., & Lau, A. H. L. (1996). The newsstand problem: a capacitated multiple-product single-period inventory problem. *European Journal of Operational Research*, *94*, 29–42.

Mahadevan, B., Pyke, D. F., & Fleischmann, M. (2003). Periodic review, push inventory policies for remanufacturing. *European Journal of Operational Research*, *151*, 536–551.

Mannella, M. (2003). What your returns are telling you. *APICS–The performance advantage*, *13*, 38–44.

Mollenkopf, D. (2006). Environmental sustainability: examining the case for environmentally-sustainable supply chains. *CSCMP Explores …*, *3*(3), 1–15.

Moon, I., & Silver, E.(2000). The multi-item newsvendor problem with a budget constraint and fixed ordering cost. *Journal of the Operational Research Society*, *51*, 602–608.

Mostard, J., & Teunter, R. (2006). The newsboy problem with resalable returns: a single-period model and case study. *European Journal of Operational Research*, *169*, 81–96.

Mukhopadhyay, S. K., Ma, H. (2009). Joint procurement and production decisions in remanufacturing under quality and demand uncertainty. *International Journal of Production Economics*, *120*, 5–17.

Niederhoff, J. (2007). Using separable programming to solve the multi-product multiple ex-ante constraint newsvendor problem and extensions. *European Journal of Operational Research*, *176*, 941–955.
Noori, A. H., & Keller, G. (1986). One-period order quantity strategy with uncertain match between the amount received and quantity requisitioned. *INFOR*, *24*, 1–11.
Rekik, Y., Sahin, E., & Dallery, Y. (2007). A comprehensive analysis of the newsvendor model with unreliable supply. *OR Spectrum*, *29*(2), 207–233.
Rogers, D. S., & Tibben-Lembke, R. S. (1999). *Going backwards: Reverse Logistics trends and practices*. Reverse Logistics Executive Council Press: Pittsburgh.
Rogers, D. S., & Tibben-Lembke, R. S. (2006). Returns management and reverse logistics for competitive advantage. *CSCMP Explores ...*, *3*(1), 1–16.
Savaskan, R. C., Bhattacharya, S., & Van Wassenhove, L. N. (2004). Closed-loop supply chain models with remanufacturing. *Management Science*, *50*(2), 239–252.
Tang, O., & Teunter, R. (2006). Economic lot scheduling problem with returns. *Production and Operations Management*, *15*(4), 488–497.
Teunter, R. H. (2001). Economic ordering quantities for recoverable item inventory systems. *Naval Research Logistics*, *48*, 484–495.
Vairaktarakis, G. L. (2000). Robust multi-item Newsboy models with a budget constraint. *International Journal of Production Economics*, *66*, 213–226.
Wyld, D. C. (2004). Reverse potential. *APICS–The performance advantage*, *14*(2), 26–32.
Yang, S., J. Yang, & Abdel-Malek, L. (2007). Sourcing with random yields and stochastic demand: a newsvendor approach. *Computers and Operations Research*, *34*, 3682–3690.
Yano, C. A., & Lee, H. L. (1995). Lot sizing with random yields: a review. *Operations Research*, *43*, 311–334.

Chapter 11
Inventory Centralization in a Newsvendor Setting When Shortage Costs Differ: Priorities and Costs Allocation

Niv Ben-Zvi and Yigal Gerchak

Abstract Risk pooling is a practical managerial tool which can reduce the consequences of the uncertainty involved in a system. In operations management, it is often achieved by consolidating a product with a random demands into one location, which is known to be beneficial. The basic assumption that underlies most previous research is that the cost parameters (overage and underage cost per unit) of all populations are identical, and therefore are equal to those of the centralized system. But in many contexts, underage cost per unit is not independent of the type of customer. This work generalizes the centralized inventory model so that one group of retailers differs from another in the underage cost per unit. In such a system, the proper allocation of the centralized inventory among the groups is a challenge. When the inventory is not allocated optimally, the expected cost of the centralized system may exceed that of the decentralized one. We define a priority rule for allocating the pooled inventory and prove that giving absolute priority to the population whose underage cost is higher ("preferred population") is optimal. Under this policy, we model the pooled inventory system with priorities and prove its advantage over the un-pooled system. We then prove the advantage of the pooled inventory system with absolute priority from each retailer's point of view, meaning that the core of the cooperative inventory game is not empty. Thus, with appropriate cost allocation, it is better to join the pool even if you were to become a low-priority customer. Finally, we introduce a pooled inventory model where the inventory is allocated according to each retailer contribution to the system, which is defined as the number of units it produces and deposits in the central warehouse. We use game theory concepts to model this system where each player's strategy is the number of units it contributes to the system. We prove the existence and uniqueness of the Nash equilibrium and characterize each player's strategy according to it.

N. Ben-Zvi • Y. Gerchak (✉)
Department of Industrial Engineering, Tel-Aviv University, Tel-Aviv 69978, Israel
e-mail: benzviniv@yahoo.com; ygerchak@eng.tau.ac.il

T.-M. Choi (ed.), *Handbook of Newsvendor Problems*, International Series in Operations Research & Management Science 176, DOI 10.1007/978-1-4614-3600-3_11,

Keywords Newsvendor problem • Inventory pooling • Asymmetric costs • Core • Nash equilibrium

11.1 Introduction

Risk pooling, or statistical economies of scale, is a pervasive property which underlies major economic activities such as banking and insurance. In the field of Operations Management, such a benefit is known to be derived from strategies like postponement (delayed differentiation), product substitutability, and assemble-to-order systems with component commonality (e,g., Gerchak and Henig 1989). But the most basic and best- known manifestation of risk pooling in operations is the centralization of inventories when demands are random. Eppen (1979), Chen and Lin (1989), Cherikh (2000), Gerchak and He (2003), and others proved the benefits of such centralization under various assumptions, and explored the determinants of these benefits. Note that pooling can be achieved by physical pooling of inventory, virtual pooling, or product substitution (e.g., Anupindi et al. 2001). For the intuition behind the benefits of pooling, see Sobel (2008). Note that the inventory level can increase or decrease as a result of pooling (Gerchak and Mossman 1992, Sect. 3, Yang and Schrage 2009).

These observations take the system (social) point of view. Yet it is plausible that the distributed system consists of several firms, which are considering as to whether centralizing their inventories will be beneficial to all of them, as they intend to remain independent regardless. That is, whether there always exists an allocation of the pooled system costs' which is lower than each firm's stand-alone costs. Müller et al. (2002) answered that question in the affirmative, by proving that the core of this cooperative game is never empty.

All these works assumed that the unit shortage cost is the same for all firms. But such may not be the case in practice. For example, if a hospital and adjacent pharmacy are to centralize their stock of some medical equipment or materials, a shortage of an item means something quite different at the hospital and the pharmacy. Differential shortage costs raise issues of stock rationing, priorities, and a question, of whether a prospective low-priority partner can be enticed to join such shared inventory arrangements. These are the issues explored here. Since our focus is on comparisons of schemes, we do not really need to estimate the shortage cost parameters (just their relative magnitudes). Thus, the thorny issue of defining and estimating shortage costs is not central to our work.

Deshpande et al. (2003) consider an inventory system with continuous rationing to demand classes. Anupindi et al. (2001) assume that each player keeps its own inventory, giving itself priority over it, and if excess occurs there is a transshipment game. That setting is related to our last, "contribution," scheme. Also related is work by Slikker et al. (2005), but they assume transferable utility (TU), while we do not. Hanany and Gerchak (2008) analyze a nontransferable utility model solved via Nash Bargaining.

Assemble-to-order systems with component commonality actually allocate components for each realization of demands, so that revenue will be maximized. That has similarity to allocating by shortage cost, but the stochastic program it translates to cannot be solved analytically, so insights are hard to obtain. Our approach is analytical.

We first show that if differential shortage costs are ignored by the centralized system when it allocates stock in case of shortage, the system costs can be higher than the decentralized systems' costs. Thus to guarantee improvement, priorities, which depend on the shortage costs, need to be set. One extreme approach, common in various service systems, is giving absolute priority (AP) to the firm with higher shortage costs. A less extreme class of policies is to allocate based on the ratios of shortage costs, either directly or according to a function of these ratios. We show that these always bring about an improvement over the decentralized system. Various properties of such systems are explored.

An interesting question is whether the low priority firm always gains from centralization, without the high priority firm subsidizing them "too much." We show that there exists a linear cost-sharing scheme which, for some range of shares, improves the situation of both firms vis-à-vis decentralized systems. That is, the core of this allocation game is not empty.

Finally, we explore a centralized system which is based on a very different principle. Here the centralized stock consists of stocks which the individual firms "contributed." In case of shortage, each firm is allocated up to its share of contribution. We characterize the Nash equilibrium of this game.

11.2 Pooling Without Recognizing Differences in Shortage Costs

Consider two markets, with random demands X and Y, with demand distributions F and G, respectively. The unit production cost c is the same in both markets. From what we know about the Newsvendor problem, if the unit shortage cost b is also the same in both markets, the total optimal inventory level will be

$$Q^*(X) + Q^*(Y) = F^{-1}\left(\frac{b-c}{b}\right) + G^{-1}\left(\frac{b-c}{b}\right), \tag{11.1}$$

and the costs

$$\phi^*(X) + \phi^*(Y) = b \int_{F^{-1}\left(\frac{b-c}{b}\right)}^{\infty} x \cdot f(x)\mathrm{d}x + b \int_{G^{-1}\left(\frac{b-c}{b}\right)}^{\infty} y \cdot g(y)\mathrm{d}y. \tag{11.2}$$

When the inventories are pooled, the combined demand, $X+Y$, will be denoted by Z, and its distribution by H. The optimal pooled quantity is $Q^*(X+Y)$. Thus, the pooled inventory is

$$Q^*(X+Y) = H^{-1}\left(\frac{b-c}{b}\right), \tag{11.3}$$

and

$$\phi^*(X+Y) = b \int_{H^{-1}\left(\frac{b-c}{b}\right)}^{\infty} z \cdot h(z)\mathrm{d}z. \tag{11.4}$$

The well-known advantage of the pooled system can be written as

$$\phi^*(X+Y) \leq \phi^*(X) + \phi^*(Y). \tag{11.5}$$

(Eppen 1979; Chen and Lin 1989; Gerchak and He 2003). This and all our other results also hold for n markets.

Now, if the unit shortage cost is market specific, we will have

$$Q_X^* + Q_Y^* = F^{-1}\left(\frac{b_X - c}{b_X}\right) + G^{-1}\left(\frac{b_Y - c}{b_Y}\right), \tag{11.6}$$

and

$$\varphi_X^* + \varphi_Y^* = b_X \cdot \int_{Q_X^*}^{\infty} x \cdot f(x)\mathrm{d}x + b_Y \cdot \int_{Q_Y^*}^{\infty} y \cdot g(y)\mathrm{d}y. \tag{11.7}$$

As the pooled system does not use any priorities, a plausible assumption is that, in case of shortage, it allocates proportionally to the relative demand realizations, x and y; that is, X receives $Q_{X+Y} \cdot \frac{x}{x+y}$, making its shortage

$$(x+y-Q_{X+Y}) \cdot \frac{x}{x+y} \tag{11.8}$$

and similarly for Y.

So the total shortage cost will be

$$b_X \cdot \left\{(x+y-Q_{X+Y}) \cdot \frac{x}{x+y}\right\} + b_Y \cdot \left\{(x+y-Q_{X+Y}) \cdot \frac{y}{x+y}\right\}. \tag{11.9}$$

Define $\bar{b} \equiv \frac{b_X \cdot x + b_Y \cdot y}{x+y}$, as the average shortage cost weighted according to the realizations. Then the total shortage costs can be written as

$$\overline{b} \cdot (x+y-Q_{X+Y}). \tag{11.10}$$

It is difficult to analyze such systema .

A simpler way to average the shortage costs is to use the mean demands rather than realized demands. That is,

$$\bar{b}_\mu = \frac{b_X \cdot \mu_X + b_Y \cdot \mu_Y}{\mu_X + \mu_Y} \,. \tag{11.11}$$

Then the problem can be solved explicitly:

$$Q^*_{X+Y} = H^{-1}\left(\frac{\bar{b}_\mu - c}{\bar{b}_\mu}\right), \tag{11.12}$$

and, as in (11.4),

$$\phi^*(X+Y) = \bar{b}_\mu \int\limits_{H^{-1}\left(\frac{\bar{b}_\mu - c}{\bar{b}_\mu}\right)}^{\infty} z \cdot h(z)\mathrm{d}z. \tag{11.13}$$

One can construct an example where the total cost of the pooled system (which ignores the differential shortage costs) is *higher* than the sum of the separate markets' costs. Suppose that both demands are independent and exponentially distributed with $\lambda_X = 0.5$ and $\lambda_Y = 0.1$, respectively. Let $b_X = 30$ and $b_Y = 2$, and $c = 1$. Then $\phi^*(X) + \phi^*(Y) = 25.734$. If the allocation is by demand realization, we have $\phi^*(X+Y) = 26.4$. If we allocate by weighting the shortage costs by the expected demands, we obtain $\phi^*(X+Y) = 31.203$. So, in this example, pooling without recognizing the differential shortage costs is not worthwhile using either method.

Not surprisingly then, we look for a system that gives priority to demands from markets with higher shortage cost.

11.3 Pooling with Priorities

Assume that $b_X > b_Y$. Then *absolute priority* (AP) allocates to market X first, and only if any inventory is left to Y. A less extreme priority scheme which reflects the differential shortage costs is one that allocated to market X a fraction $\frac{b_X}{b_X+b_Y}$ of available inventory (if needed); the rest is available to market Y. Thus if, say, b_X is twice as large as b_Y, market X will be allocated $^2/_3$. Let αp be "Alpha-priority" (see below).

One way to parameterize allocation schemes which are intermediate between the above two is

$$\left(\frac{b_X}{b_X+b_Y}\right)^{\alpha} \cdot Q_{\alpha p}(X+Y), \text{ to } X, \text{ and } \left[1 - \left(\frac{b_X}{b_X+b_Y}\right)^{\alpha}\right] \cdot Q_{\alpha p}(X+Y) \text{ to } Y, \tag{11.14}$$

where α is a "priority" parameter on [0,1]. The larger the α, the lower the priority. It can be proved that $\alpha = 0$, i.e., an absolute priority to X, is socially optimal.

We will thus analyze a system with absolute priority[1].

11.3.1 Absolute Priority

Under AP, if $x \leq Q_{AP}(X+Y)$, market X will have no shortage, while market Y's shortage will be $(y-(Q_{AP}(X+Y)-x))^{+}$. If $x > Q(X+Y)$, market X will have a shortage $(x-Q(X+Y))$, and market Y will have a shortage y. Thus,

$$\begin{aligned}
&\phi_{X+Y}(Q)\\
&= \min_{Q_{X+Y}} \left\{ \begin{array}{c} c \cdot Q_{X+Y} + b_Y \cdot E_{X,Y}\left[\left(Y-\left(Q_{X+Y}-X\right)\right)|X \leq Q_{X+Y}, Y \geq Q_{X+Y}-x\right] \\ +b_Y \cdot P\left(X \geq Q_{X+Y}\right) \cdot E_Y\left(Y\right) + b_X \cdot E_X\left(X-Q_{X+Y}\right)^{+} \end{array} \right\}\\
&= \min_{Q_{X+Y}} \left\{ c \cdot Q_{X+Y} + b_Y \cdot \iint\limits_{\substack{x \leq Q_{X+Y}\\ y \geq Q_{X+Y}-x}} \left(y-\left(Q_{X+Y}-x\right)\right) \cdot t\left(x,y\right) \mathrm{d}x\mathrm{d}y \right.\\
&\qquad \left. + b_Y \cdot \overline{F}\left(Q_{X+Y}\right) \cdot E\left(Y\right) + b_X \cdot \int\limits_{Q_{X+Y}}^{\infty} \left(x-Q_{X+Y}\right) \cdot f(x)\mathrm{d}x \right\}, \qquad (11.15)
\end{aligned}$$

where $t(x,y)$ is the joint density of X and Y.

For independent demands, the above becomes

$$\begin{aligned}
&\phi_{X+Y}\left(Q_{X+Y}\right)\\
&= \min_{Q_{X+Y}} \left\{ c \cdot Q_{X+Y} + b_Y \cdot \int\limits_{x=0}^{Q_{X+Y}} \int\limits_{y=Q_{X+Y}-x}^{\infty} \left(y-\left(Q_{X+Y}-x\right)\right) \cdot g(y) \cdot f(x)\mathrm{d}y\mathrm{d}x \right.\\
&\qquad \left. + b_Y \cdot \overline{F}\left(Q_{X+Y}\right) \cdot E\left(Y\right) + b_X \cdot \int\limits_{Q_{X+Y}}^{\infty} \left(x-Q_{X+Y}\right) \cdot f(x)\mathrm{d}x \right\}. \qquad (11.16)
\end{aligned}$$

[1]For the expected costs and optimality condition expression for an arbitrary α and independent demands, see Appendix.

Since

$$\frac{\partial^2 \phi_{X+Y}(Q)}{\partial Q_{X+Y}^2} = (b_X - b_Y) \cdot f(Q_{X+Y}) + b_Y \cdot \int_{x=0}^{Q_{X+Y}} g(Q_{X+Y} - x) \cdot f(x) \mathrm{d}x,$$

as $b_X > b_Y$ that is positive so the expected cost function is convex. The unique optimum is the solution of the FOC

$$(b_X - b_Y) \cdot F_X(Q_{X+Y}) + b_Y \cdot H(Q_{X+Y}) = b_X - c. \tag{11.17}$$

This optimality condition can be shown to generalize the simpler cases we mentioned. If $b_X = b_Y \equiv b$, we have $H(Q_{X+Y}) = P(X+Y \leq Q_{X+Y}) = \frac{b-c}{b}$, which is the pooled solution with identical shortage costs. If $b_Y \cong 0$, we have $F(Q_{X+Y}) = \frac{b_X - c}{b_X}$, as in a single-market newsvendor solution. One can show that the quantity in the pooled system is increasing in both b_X and b_Y, as we would intuitively expect.

Under AP, pooling is beneficial:

Proposition 1. *Under AP*

$$\phi_{X+Y}^*(AP) \leq \phi_X^* + \phi_Y^*.$$

Proof. See Appendix.

Also, in a pooled system AP is always better than no priority (NP), where the shortage cost is weighed by the realized values.

Proposition 2.

$$\phi_{X+Y}^*(AP) \leq \phi_{X+Y}^*(NP).$$

Proof. See Appendix.

11.4 Do Both Markets Benefit from Pooling with AP?

While pooling with AP was shown to be beneficial socially, it is not yet clear if or when the low-priority market benefits from the arrangement. If market X approaches market Y proposing to pool their inventories whereby X will have AP, should Y agree?

For pooling with equal shortage costs it has been proved by Müller et al. (2002) that the *core* of such a game is not empty. That is, there is a way to allocate the costs of the pooled system so that both markets will benefit, in agreement with intuition. With AP to the more lucrative market, the situation is less clear.

We now assume that each market pays for a fixed share of the production costs, Y a share θ and X a share $1-\theta$. The reason we allocate only the production costs is that, unlike the shortage costs, they will be clear-cut and noncontroversial. But each market will also "carry" its shortage costs. We will then ask if there always is a range of shares which makes pooling with AP better, in expectation, for both markets than nonpooled systems. Note that θ does not affect the solution Q^*_{X+Y} of the pooled system. Denoting the cost allocated to market X as $\phi_{X+Y}(X)$, we have

$$\phi^*_{X+Y}(X) = (1-\theta)\cdot c\cdot Q^*_{X+Y} + b_X \cdot \int\limits_{Q^*_{X+Y}}^{\infty} (x - Q^*_{X+Y})\cdot f(x)\mathrm{d}x, \tag{11.18}$$

while in the unpooled system its costs are

$$\varphi^*_X(X) = c\cdot Q^*_X + b_X \cdot \int\limits_{Q^*_X}^{\infty} (x - Q^*_X)\cdot f(x)\mathrm{d}x. \tag{11.19}$$

Clearly, if the low priority market's share θ is sufficiently high, the arrangement will be attractive to the high priority market. In particular, $\phi^*_{X+Y}(X) \le \phi^*_X(X)$. If θ is sufficiently low, the high priority market will not participate in the pooling. Let $\theta_{\min}$ be the break-even point. The situation is reversed for the low priority market. Now, one can show that for $\theta = \theta_{\min}$ (the breakeven point of the high priority market) $\phi^*_Y(Y) \ge \phi^*_{X+Y}(Y)$. Thus, the core of this game is not empty. In fact, even if θ is slightly higher than $\theta_{\min}$ the low priority market will still benefit. Thus, the core is a range of shares from $\theta_{\min}$ upward to some point.

We are unable to say much about the upper end of the core in the general model. To obtain some concrete insight, we derive it for uniformly distributed independent demands.

Example 1. Let X and Y be independent and uniform on [0,1]. It can be shown that their convolution is

$$H(Q) = \begin{cases} -\frac{1}{2}Q^2 + 2Q - 1 & 1 \le Q \le 2 \\ \frac{1}{2}Q^2 & Q < 1 \end{cases}$$

and thus

$$Q^* = \begin{cases} 2 - \sqrt{\frac{2c}{b_Y}}, & 1 \le Q \le 2, \\ \frac{b_Y - b_X + \sqrt{b_X^2 + b_Y^2 - 2b_Y c}}{b_Y}, & Q < 1. \end{cases}$$

Also $Q < 1$ iff $\frac{2c}{b_Y} > 1$. If $\frac{2c}{b_Y} < 1$, then $\phi^*_{X+Y} = 2 - \sqrt{\frac{2c}{b_Y}} > 1$.

Since $X \le 1$, then $\bar{F}(Q^*_{X+Y}) = 0$. Thus, the third term in (11.16), as well as the fourth, equal zero.

Thus, here

$$\varphi(X+Y) = c\cdot Q^* + b_Y \cdot \int\limits_{x=0}^{Q^*} \int\limits_{y=Q^*-x}^{\infty} (y - (Q^* - x))\cdot g(y)\cdot f(x)\mathrm{d}y\mathrm{d}x,$$

and thus it can be shown that at optimum for $\frac{2c}{b_Y} < 1$, we have

$$\phi^*_{X+Y} = 2c\left(1 - \frac{1}{3}\sqrt{\frac{2c}{b_Y}}\right).$$

The case $\frac{2c}{b_Y} > 1$ is more complex, and will not be pursued.

When the markets operate separately, we have, for our example,

$$\phi^*_Y = c\left(1 - \frac{1}{2} \cdot \frac{c}{b_Y}\right)$$

and

$$\phi^*_Y = c\left(1 - \frac{1}{2} \cdot \frac{c}{b_X}\right).$$

Consider the case $\frac{2c}{b_Y} < 1$. The costs share of the low priority market is

$$\phi^*_{X+Y}(Y) = \theta \cdot c \cdot \left(2 - \sqrt{\frac{2 \cdot c}{b_Y}}\right) + \left(\frac{1}{3} \cdot c \cdot \sqrt{\frac{2 \cdot c}{b_Y}}\right).$$

It follows that Y will benefit from pooling if

$$\theta < \frac{c \cdot \left(1 - \frac{1}{2} \cdot \frac{c}{b_Y}\right) - \left(\frac{1}{3} \cdot c \cdot \sqrt{\frac{2 \cdot c}{b_Y}}\right)}{c \cdot \left(2 - \sqrt{\frac{2 \cdot c}{b_Y}}\right)}$$

$$= \frac{1 - \frac{1}{2} \cdot \frac{c}{b_Y} - \frac{1}{3} \cdot \sqrt{\frac{2 \cdot c}{b_Y}}}{2 - \sqrt{\frac{2 \cdot c}{b_Y}}}.$$

Market X's share in the costs is

$$\phi^*_{X+Y}(X) = (1 - \theta) \cdot c \cdot \left(2 - \sqrt{\frac{2 \cdot c}{b_Y}}\right),$$

which means that it will benefit if

$$\theta > \frac{2 - \sqrt{\frac{2 \cdot c}{b_Y}} - 1 + \frac{1}{2} \cdot \frac{c}{b_X}}{2 - \sqrt{\frac{2 \cdot c}{b_Y}}}.$$

The lower bound on θ can be shown to be indeed smaller than the upper bound.

Thus, in this case the core is

$$0 < \frac{1-\sqrt{\frac{2\cdot c}{b_Y}}+\frac{1}{2}\cdot\frac{c}{b_X}}{2-\sqrt{\frac{2\cdot c}{b_Y}}} < \theta < \frac{1-\frac{1}{2}\cdot\frac{c}{b_Y}-\frac{1}{3}\cdot\sqrt{\frac{2\cdot c}{b_Y}}}{2-\sqrt{\frac{2\cdot c}{b_Y}}} < 1.$$

This core is not empty.

11.5 Allocation According to Contribution to Inventory

Here we view the system in a different, and rather unusual, manner. We envision a centralized inventory which is created by the individual firms. Each decides on the quantity it will contribute to the joint stock. The decisions are simultaneous. Then, if there is a shortage, each will receive an allocation proportional to its relative contribution. There are no imposed priorities. Thus we have a noncooperative game, somewhat related to the model of decentralized transshipment by Rudi et al. (2001), and Hanany et al. (2010), and we seek its Nash equilibrium.

Assume that the firms produce/contribute Q_X and Q_Y, respectively. In case of shortage, firm X is allocated $\frac{Q_X}{Q_X+Q_Y}$ of the stock, and firm Y, $\frac{Q_Y}{Q_X+Q_Y}$. Essentially, each firm is first allocated its contribution, or as much as it requires, whichever is less, and then any excess from another firm, if it exists and is needed. The firms are aware of that when they choose their contributions.

Given Q_Y, firms X's cost function is

$$\begin{aligned}\phi_X(Q_X,Q_Y) &= cQ_X + b_X E\left\{(X-Q_X)^+\right\}P(Y>Q_Y)\\ &\quad + b_X E\left\{(X+Y-Q_X-Q_Y)^+ \,|X>Q_X\,,Y<Q_Y\right\}.\end{aligned} \tag{11.20}$$

Assuming independence that can be shown to equal

$$\begin{aligned}\phi_X(Q_X,Q_Y) &= c\cdot Q_X + b_X\cdot\overline{G}(Q_Y)\cdot\int_{x=Q_X}^{\infty}(x-Q_X)\cdot f(x)\mathrm{d}x\\ &\quad + b_X\cdot\int_{y=0}^{Q_Y}\int_{x=Q_X+Q_Y-y}^{\infty}(x+y-Q_X-Q_Y)\cdot f(x)\cdot g(y)\mathrm{d}x\mathrm{d}y,\end{aligned} \tag{11.21}$$

and similarly for ϕ_Y, given Q_X.

Differentiating each cost function (which is convex) with respect to its strategy, we obtain the best-response functions

$$(i)\qquad c-b_X\cdot\overline{G}(Q_Y^*)\cdot\overline{F}(Q_X)-b_X\cdot\int_{y=0}^{Q_Y^*}\overline{F}(Q_X+Q_Y^*-y)\cdot g(y)\mathrm{d}y=0 \tag{11.22}$$

$$(ii) \qquad c - b_Y \cdot \overline{F}\left(Q_X^*\right) \cdot \overline{G}\left(Q_Y\right) - b_Y \cdot \int_{x=0}^{Q_X^*} \overline{G}\left(Q_X^* + Q_Y - x\right) \cdot f(x)\mathrm{d}x = 0\,. \qquad (11.23)$$

The Nash equilibrium obtained by solving these equations can be shown to exist and is unique. We shall prove uniqueness:

We use a contraction-mapping argument (e.g., Cachon and Netessine 2004). Now,

$$\frac{\partial Q_X^*}{\partial Q_Y} = -\frac{\frac{\partial^2 \phi_X}{\partial Q_X \partial Q_Y}}{\frac{\partial^2 \phi_X}{\partial Q_X^2}} = -\frac{b_X \cdot \int_{y=0}^{Q_Y} f\left(Q_X + Q_Y - y\right) \cdot g(y)\mathrm{d}y}{b_X \cdot \overline{G}\left(Q_Y\right) \cdot f\left(Q_X\right) + b_X \cdot \int_{y=0}^{Q_Y} f\left(Q_X + Q_Y - y\right) \cdot g(y)\mathrm{d}y} < 0\,, \qquad (11.24)$$

which is not surprising, as there is a "free rider" situation here.

Also,

$$\left|\frac{\partial Q_X^*}{\partial Q_Y}\right| < 1 \text{ and } \left|\frac{\partial Q_Y^*}{\partial Q_X}\right| < 1. \qquad (11.25)$$

That establishes that the response functions are contractions.

It can be shown that $\frac{\mathrm{d}Q_X}{\mathrm{d}b_X} > 0$, which is intuitive, and $\frac{\mathrm{d}Q_Y}{\mathrm{d}b_X} < 0$, which occurs since firm Y knows that with high shortage cost firm X will contribute a large amount, so Y itself need not contribute much.

One can show that the amounts contributed in the Nash equilibrium are lower than in the decentralized system. That is consistent with observations made by others about inventory games with substitutable products.

11.6 Concluding Remarks

Risk pooling underlies many types of arrangements in the financial sector (banks, insurance companies) as well as in various operational settings. The benefits in some of these settings were explored before, but mainly for systems where customers are treated equally. But customers are not always, nor should they be from a social perspective, treated equally. The waiting costs of one population may be higher than the other's. Or, one class of customers has more power. In such settings an important issue is whether, or when, the would-be low priority customer class will still benefit from the pooled arrangement,. That is, whether the core is nonempty. These and related issues were explored here. We also considered a setting where the customer classes choose their "contributions" to the pooled inventory, and, in case of shortage, are allocated units in proportion to their relative contribution. We establish the existence and uniqueness of a Nash equilibrium in that model and characterize it. We wish to note that the setting-up of schemes like AP, or contribution-based allocation, do not require the knowledge of actual shortage costs. Thus, the conceptually and practically difficult task of estimating these nebulous costs is not encountered at this phase.

Table 11.1 Shortage allocation

	Shortage				
	Centralized system with absolute priority			Decentralized system	
	X	Y		X	Y
$x \leq Q_X$ $y \leq Q_Y$	0	0	$\leq$	0	0
$x \leq Q_X$ $y \geq Q_Y$ $x+y \leq Q_X+Q_Y$	0	0	$\leq$	0	$y-Q_Y$
$x \leq Q_X$ $y \geq Q_Y$ $x+y \geq Q_X+Q_Y$	0	$y-(Q_X+Q_Y-x) \leq y-Q_Y$	$\leq$	0	$y-Q_Y$
$x \geq Q_X$ $y \geq Q_Y$ $x \leq Q_X+Q_Y$	0	$y-(Q_X+Q_Y-x)$	$\leq$	$x-Q_X$	$y-Q_Y$
$x \geq Q_X$ $y \geq Q_Y$ $x \geq Q_X+Q_Y$	$x-(Q_X+Q_Y)$	y	$\leq$	$x-Q_X$	$y-Q_Y$
$x \geq Q_X$ $y \leq Q_Y$ $x+y \leq Q_X+Q_Y$ $\Rightarrow x \leq Q_X+Q_Y$	0	0	$\leq$	$x-Q_X$	0
$x \geq Q_X$ $y \leq Q_Y$ $x+y \geq Q_X+Q_Y$ $x \leq Q_X+Q_Y$	0	$y-(Q_X+Q_Y-x) \leq x-Q_X$	$\leq$	$x-Q_X$	0
$x \geq Q_X$ $y \leq Q_Y$ $x \geq Q_X+Q_Y$ $\Rightarrow x+y \geq Q_X+Q_Y$	$x-(Q_X+Q_Y)$	y	$\leq$	$x-Q_X$	0

Further research could address the extension from two to n populations. The notion of priority needs to be expanded to deal with that case. Dependencies among group demand magnitudes also should be explored. A multi-period model with inventory carry over would also be of interest. A different direction is to derive entitlements (as well as stocks) in the form of Nash Barganing Solutions (Hanany and Gerchak 2008).

The model which assumes that the parties make inventory contributions, ought to be explored further. The existence of a nonempty core will be of a particular interest.

Appendix

Proof of Proposition 1. Let Q_X and Q_Y be the optimal respective quantities in the decentralized system. Now suppose that the centralized system will hold a quantity of $Q_X + Q_Y$ (that is not optimal, but if we show that a centralized system holding that amount has lower costs than the decentralized system, the optimal centralized system costs will be even lower). We now show that for a demand realization (x, y)

Table 11.2 AP vs. NP

Realizations	Expected shortage cost with AP	Expected shortage costs with NP
$x+y \leq Q_{NP}^{*}$	0	0
$x+y \geq Q_{NP}^{*}$ $x \leq Q_{NP}^{*}$	$b_X \cdot 0 + b_Y \cdot (x+y-Q_{NP}^{*}) = b_Y \cdot (x+y-Q_{NP}^{*})$	$\overline{b} \cdot (x+y-Q_{NP}^{*})$
$x+y \geq Q_{NP}^{*}$ $x \geq Q_{NP}^{*}$	$b_X \cdot (x-Q_{NP}^{*}) + b_Y \cdot y$	$\overline{b} \cdot (x+y-Q_{NP}^{*})$

the magnitude of shortage in the centralized system will be smaller than in the decentralized for each of the two groups (Table 11.1).

Now, since production costs are equal and the shortage of population X as well as the total shortage are lower, it follows that the costs of the centralized system with AP are lower than the decentralized system. As that is the case for a nonoptimal centralized stock, it is a fortiori the case for the optimal stock.

Proof of Proposition 2. Let Q_{NP}^{*} be the optimal quantity in the no-priority system, and assume that the AP system stocks the same amount (which is not optimal) (Table 11.2). Then the expected shortage costs of the two systems for various demands realizations are

In the second case, since $\bar{b} \geq b_Y$ the expected shortage with AP is clearly lower. In the third case, the difference between the expected costs with AP and NP is

$$(b_X \cdot (x-Q_{NP}^{*}) + b_Y \cdot y) - (\overline{b} \cdot (x+y-Q_{NP}^{*})) = \underbrace{\frac{y}{x+y} \cdot Q_{NP}^{*}}_{>0} \cdot \underbrace{(b_Y - b_X)}_{<0},$$

which is clearly negative. Thus then expected costs with AP is lower than with NP for all demand realizations.

Arbitrary Priority

It can be shown that for independent demands, and denoting $\beta = b_X/(b_X+b_Y)$, we obtain the following expected costs function:

$$j^{*}(X+Y) = \min \left\{ \begin{array}{l} c \times Q_{X+Y} + b_X \times \int\limits_{y=0}^{(1-\beta^{\alpha})Q_{X+Y}} \int\limits_{x=Q_{X+Y}-y}^{¥} (x+y-Q_{X+Y}) \times f(x) \times g(y)\mathrm{d}x\mathrm{d}y \\ +b_X \times \bar{G}((1-\beta^{\alpha}) \times Q_{X+Y}) \times \int\limits_{x=\beta^{\alpha}Q_{X+Y}}^{¥} (x-\beta^{\alpha} \times Q_{X+Y}) \times f(x)\mathrm{d}x \\ +b_Y \times \int\limits_{x=0}^{\beta^{\alpha}Q_{X+Y}} \int\limits_{y=Q_{X+Y}-x}^{¥} (x+y-Q_{X+Y}) \times f(x) \times g(x)\mathrm{d}x\mathrm{d}y \\ +b_Y \times \bar{F}(\beta^{\alpha} \times Q_{X+Y}) \times \int\limits_{y=(1-\beta^{\alpha})Q_{X+Y}}^{¥} (y-(1-\beta^{\alpha}) \times Q_{X+Y}) \times g(y)\mathrm{d}y \end{array} \right\}.$$

The optimality equation becomes

$$c - b_X \cdot \int_{y=0}^{(1-\beta^{\alpha})\cdot Q_{X+Y}} \bar{F}\left(Q_{X+Y} - y\right) \cdot g(y)\mathrm{d}y - b_Y \cdot \int_{x=0}^{\beta^{\alpha}\cdot Q_{X+Y}} \bar{G}\left(Q_{X+Y} - x\right) \cdot f(x)\mathrm{d}x$$
$$-\bar{F}\left(\beta^{\alpha} \cdot Q_{X+Y}\right) \cdot \bar{G}\left(1-\beta^{\alpha}\right) \cdot Q_{x+y} \cdot \left(b_X \cdot \beta^{\alpha} + b_Y \cdot \left(1-\beta^{\alpha}\right)\right) = 0.$$

References

Anupindi, R., Bassok, Y., & Zemel, E. (2001). A general fromework for the study of decentralized distribution systems *Manufacturing & Service Operations Management, 3*, 349–368.

Cachon, P. G., & Netessine, S. (2004). Game theory in supply chain analysis. In D. Simchi-Levi, S. D. Wu & M. Shen (Eds.), *Supply chain analysis in the eBusiness era*. Dordecht: Kluwer.

Chen, M. S., & Lin, C. T. (1989). Effects of centralization on expected costs in a multi-location newsboy problem. *Journal of Operational Research Society, 40*, 597–602.

Cherikh, M. (2000). On the effect of centralization on expected profits in a multi-location newsboy problem. *Journal of the Operational Research Society, 51*, 755–761.

Deshpande, V., Cohen, M. A., & Donohue, K. (2003). The threshold inventory rationing policy for service-differentiated demand classes. *Management Science, 49*, 683–703.

Eppen, G. D. (1979). Effects of centralization on expected costs in a multi-location newsboy problem. *Management Science, 25*, 498–501.

Gerchak, Y., & Henig, M. (1989). Component commonality in ATO systems: Models and properties. *Naval Research Logistics, 36*, 61–68.

Gerchak, Y., & He, Q. M. (2003). On the relation between the benefits of risk pooling and the variability of demand. *IIE Transaction, 35*, 1–5.

Gerchak, Y., & Mossman, D. (1992). On the effect of demand randomness on inventories and costs. *Operations Research, 40*, 804–807.

Granot, D., & Sošić, G. (2003). A three-stage model for a decentralized distribution system of retailers. *Operations Research, 51*, 771–784.

Hanany, E., & Gerchak, Y. (2008). Nash bargaining over inventory and pooling contracts. *Naval Research Logistics, 55*, 541–550.

Hanany, E., Tzur, M., & Levran, A. (2010). The transshipment fund mechanism: Coordinating the decentralized multi-location transshipment problem, *Naval Research Logistics, 57*, 342–353.

Hartman, B. C., & Dror, M. (2005). Allocation of gains from inventory centralization in newsvendor environments. *IIE Transactions, 37*, 93–107.

Klijn, F., & Slikker, M. (2005). Distribution center consolidation games. *Operations Research L etters, 33*, 285–288.

Müller, A., Scarsini, M., & Shaked, M. (2002). The newsvendor game has a nonempty core. *Games and Economics Behavior, 38*, 118–126.

Rudi, N., Kapur, S., & Pyke, D. (2001). The two location inventory model with transshipment and local decision making. *Management Science, 47*, 1668–1680.

Slikker, M., Fransoo, J., & Wouters, M. (2005). Cooperation between multiple newsvendors with transshipment. *European Journal of Operational Research, 167*(2), 370–380.

Sobel, M. J. (2008). Risk pooling. In D. Chhajed & T. S. Lowe (Eds.), *Building Intuition* (Ch. 9, pp. 155–174). Berlin: Springer.

Yang, H., & Schrage, L. (2009). Conditions that cause risk pooling to increase inventory, *European Journal of Operational Research, 192*, 837–851.

Chapter 12
Planning Production on an Unreliable Machine for Multiple Items Subject to Stochastic Demand

David Kletter

Abstract We develop an extension of the classical newsvendor model that incorporates multiple items, setup times, and an unreliable machine. This model is motivated by applications at metal stamping plants where machine reliability is a key source of uncertainty. Given a fixed production schedule, a finite horizon, and a known demand distribution, we formulate an extension of the newsvendor model, derive important properties of this model, and exploit these properties to provide a solution algorithm that determines the cost minimizing production quantities. Finally, we present three simple extensions to the model: (1) a method for rescheduling within the planning horizon, (2) an extension to evaluate whether or not to purchase the option to run overtime within the planning horizon, and (3) an extension that permits the modeling of a machine that operates at a different speed depending on the part being produced.

Keywords Multiple items • Setup times • Unreliable machines • Cost minimization • Solution algorithm • Rescheduling

12.1 Introduction

In this chapter, we develop extension of the classical newsvendor model. We model a single, unreliable machine that repetitively produces a set of parts in batches subject to shortage and overage (inventory-holding) costs. Our model makes the following assumptions. First, we assume that there is only a single demand point for all parts, and that it occurs at the end of a finite production horizon. Second, the demand for each part is a random variable with a known distribution, where the uncertainty in the demand quantity is not resolved until the demand point.

D. Kletter (✉)
Booz Allen Hamilton, New York, NY, USA
e-mail: kletter_david@bah.com

T.-M. Choi (ed.), *Handbook of Newsvendor Problems*, International Series in Operations Research & Management Science 176, DOI 10.1007/978-1-4614-3600-3_12,

Lastly, we assume a fixed production sequence. Under these assumptions, our model determines an optimal production quantity for each part. The development of this model is motivated by applications at metal stamping plants (Kletter 1994). This model could be used as part of a manufacturing control system, embedded in a software tool that would receive data in real time from the shop floor and assist plant management in decision making.

This chapter is structured as follows. Section 2 presents a review of the literature. A model is then formulated in Sect. 3 as an extension of a classical newsvendor problem. In Sect. 4, we derive properties of the objective function that are exploited to develop a solution algorithm, presented in Sect. 5 and that takes advantage of the special structure of the model. In Sect. 6, we show numerical results from exercise of the model. Finally, in Sect. 7, three extensions to the model are presented, including the incorporation of options to run overtime.

12.2 Literature Review

We briefly review the literature that is related to our model. We will divide our literature review into two parts: those that model the problem of planning production quantities on an unreliable machine, and those that use a newsvendor model for problems closely related to the one we study here.

12.2.1 Unreliable Machine

The presence of machine unreliability in a manufacturing system has been studied in a variety of different contexts, including problems of sequencing, scheduling, and lot sizing. We briefly review each of these areas.

We first discuss sequencing of jobs on an unreliable machine. The earliest work is that of Glazebrook (1984) who models the problem as a rather general cost-discounted Markov decision process. He shows the conditions under which the optimal policy is of an index type (i.e., the job to be processed is the one with the smallest *Gittins index*; see Gittins 1979). Pinedo and Rammouz (1988) find the optimal nonpreemptive policies for several objective functions in the case of a Poisson failure process. For a general failure process and a discrete time model, Birge and Glazebrook (1988) find bounds on the error of following the strategy that is optimal when the failure process is memoryless. Birge et al. (1990) study in greater detail the problem of minimizing weighted flow-time and obtain results that are consistent with and complementary to Pinedo and Rammouz. Epstein et al. (2010) analyze optimal sequencing on an unreliable machine where the machine may slow or stop completely. For a detailed and current overview of this research area, see the surveys contained in Lee (2004), Pinedo (2008), Diedrich et al. (2009), and Racke (2009).

There is also a significant body of work on lot sizing on an unreliable machine. Yano and Lee (1995) provide a broad review of the literature in this area.

See Schmidt (2000) for an overview of the scheduling literature in cases where the machine is continuously available for processing, for example, when there is incomplete information about when the machine may change availability. Al-Salamah and Abudari (2011) model a production process with failures: however, failures don't result in the stoppage of the machine but instead produces nonconforming items. El-Ferik (2008) determines optimal production quantities under an assumption that the production facility is subject to random failure, where preventative maintenance schedules need to be balanced with production. Halim et al. (2010) show the effect on optimal lot sizing when an unreliable machine is subject to fuzzy demand and repair time. Giri et al. (2005) model a two-stage production system where the upstream stage is subject to failures but the downstream stage is not. They assume that after a machine failure, production of the affected lot is not resumed.

In the case of an infinite planning horizon, Giri and Dohi (2004) derive optimal lot sizes under a net present value (NPV) approach.

Of particular note is the work of Groenevelt et al. (1992a, 1992b), who extend the basic economic manufacturing quantity (EMQ) model to incorporate the effects of machine breakdowns. The first paper assumes that repairs are instantaneous but bear a fixed cost. The second paper assumes (as we do) that repairs are not instantaneous but instead consume machine time. This model permits any repair time distribution, but assumes that the time between failures is exponentially distributed. Under the assumption of lost sales, the authors seek an optimal lot size and safety stock level to minimize cost subject to a constraint on the service level. They require, however, some awkward assumptions regarding safety stock to achieve separability in the optimization of the lot sizes and safety stock level. The authors do not explore the impacts of multiple parts sharing the same machine.

Other authors, such as Sethi and Zhang (1994) have approached similar problems from a control theoretic perspective. These authors consider the problem of finding an optimal setup schedule (a sequence of parts and the times at which the changeovers will occur) for an unreliable machine. They show that in the limit (as the length of the horizon tends to infinity), the stochastic problem can be reduced to a deterministic problem, and show how to obtain the optimal control policy. The authors also cite many other similar works.

Reiman and Wein (1998) study a two customer class, single server system with setups. The authors use heavy traffic diffusion approximations to analyze a system with a renewal arrival process, general service times, and either setup costs or setup times. They solve a control problem to minimize a linear function of the queue length plus setup costs, if any. Within these heavy traffic diffusion approximations, one could model the unreliability of the machine within the service time distribution.

12.2.2 Related Newsvendor Models

The classic "newsvendor" model has been the subject of many extensions that are similar to those we consider here. For many decades, researchers have considered models with multiple items (Evans 1967, Smith et al. 1980). Rose (1992) considers uncertain replenishment, but assumes demand is deterministic. Others have studied multiple time periods for production (Bitran et al. 1986; Matsuo 1990; Ciarallo et al. 1994). Jain and Silver (1995) model uncertainty in supply and permit the option to reserve reliable capacity for a premium charge. Dada et al. (2007) consider a newsvendor that purchases a single item provided by multiple suppliers, some of which are unreliable, and develop a model for optimal supplier selection. Huggins and Olsen (2010) extend the basic newsvendor problem for a single item to permit expediting for unmet demand.

12.3 Formulation

The mathematical structure of our model will closely parallel that of the classic newsboy model, which we now briefly describe. The following table (Table 12.1) lists the notation that we will use throughout this section.

For simplicity we will first state the formulation as a single part formulation, dropping the index i from our notation. The problem is then to choose an order-up-to quantity y to minimize the expected purchase, holding, and shortage costs. Mathematically, we can state the problem as

$$C^*(x) = \min_{y \geq x} c(y-x) + p\int_{y}^{\infty}(t-y)g(t)\mathrm{d}t + h\int_{0}^{y}(y-t)g(t)\mathrm{d}t. \tag{12.1}$$

Table 12.1 Notation for formulation

i	Index which denotes different parts to be produced; $i = 1,\ldots,N$
y_i	Decision variable denoting order-up-to level for part i
x_i	Current inventory level of part i
c_i	Unit purchase price of part i
h_i	Cost per unit of inventory remaining at the end of the period for part i
p_i	Unit shortage cost for part i
s_i	Time required to set up the machine to begin producing part i
$g_i(\cdot)$	PDF of demand for part i
$f_i(t;T)$	PDF that in T units of time, the cumulative output of the machine is t units of part i
$F_i(\cdot)$, $G_i(\cdot)$	CDFs for the PDFs f_i, g_i
$\overline{F}_i(\cdot)$, $\bar{G}_i(\cdot)$	$1-F_i(\cdot)$, $1-G_i(\cdot)$

The problem is solved by finding the value of y such that $\partial C(x)/\partial y$ is zero. To find this partial derivative, we need to employ Leibnitz's theorem for differentiation of an integral:

$$\frac{\partial}{\partial y}\int_{p(y)}^{q(y)} f(x,y)\mathrm{d}x = \int_{p(y)}^{q(y)} \frac{\partial f(x,y)}{\partial y}\mathrm{d}x + \frac{\partial q(y)}{\partial y} f(q(y),y) - \frac{\partial p(y)}{\partial y} f(p(y),y) \quad (12.2)$$

(Beyer 1987). We will use this extensively in our analysis. From this rule, it is easy to see that the optimal solution y^* to the newsboy model occurs at the point where $G(y^*) = (p-c)/(p+h)$, unless this implies $y^* < x$, in which case it is optimal not to order.

We now extend this basic single part model to our multiple part, unreliable production model, for now ignoring overtime opportunities. The problem is to find the optimal order-up-to levels to minimize the sum of purchasing, holding, and shortage costs over all parts. Let y, x, c, p, h, and $g(\cdot)$ retain the same meanings as above, except now we add a subscript i, for each part $i = 1,\ldots,N$. We assume without loss of generality that the parts are indexed in the order in which they will be produced. In practice, the order-up-to strategy would be implemented as follows: produce the first part until the inventory level reaches the optimal y_1, then the production switches to the part 2 until its inventory level reaches the optimal y_2, etc.

Let T denote the amount of time available for production, and the time available after setups as $T_i = T - s_1 - \cdots - s_i$. If we are already setup to produce part 1, then we set $s_1 = 0$. We assume for simplicity that each part is produced at the same rate when the machine is working. We now introduce machine unreliability into the formulation by including the PDF $f(t;T)$. Kletter (1996) provides a variety of formulas for this distribution in the case where the interarrival time of machine failures and repairs are exponentially distributed. However, the results below are independent of the form of this distribution.

We can now write the problem as

$$C^*(x) = \min_{y_1 \geq x_1,\ldots,y_N \geq x_N} C(y,x), \quad (12.3)$$

where

$$C(y,x) = \sum_{i=1}^{N} C_i(y,x), \quad (12.4)$$

and

$$C_i(y,x) = c_i \int_{x_i}^{y_i} (t-x_i) f\left(\sum_{j=1}^{i-1} y_j - x_j + t - x_i; T_i\right) \mathrm{d}t + c_i(y_i - x_i)\bar{F}\left(\sum_{j=1}^{i} y_j - x_j; T_i\right)$$
$$+ p_i \int_{x_i}^{y_i}\int_{u}^{\infty} (t-u) g_i(t) \mathrm{d}t\, f\left(\sum_{j=1}^{i-1} y_j - x_j + u - x_i; T_i\right) \mathrm{d}u$$

$$+p_i\bar{F}\left(\sum_{j=1}^{i} y_j - x_j; T_i\right)\int_{y_i}^{\infty}(t-y_i)g_i(t)\mathrm{d}t$$

$$+p_iF\left(\sum_{j=1}^{i-1} y_j - x_j; T_i\right)\int_{x_i}^{\infty}(t-x_i)g_i(t)\mathrm{d}t$$

$$+h_i\int_{x_i}^{y_i}\int_{0}^{u}(u-t)g_i(t)\mathrm{d}t f\left(\sum_{j=1}^{i-1} y_j - x_j + u - x_i; T_i\right)\mathrm{d}u$$

$$+h_i\bar{F}\left(\sum_{j=1}^{i} y_j - x_j; T_i\right)\int_{0}^{y_i}(y_i-t)g_i(t)\mathrm{d}t$$

$$+h_iF\left(\sum_{j=1}^{i-1} y_j - x_j; T_i\right)\int_{0}^{x_i}(x_i-t)g_i(t)\mathrm{d}t, \tag{12.5}$$

where the summations from 1 to $i-1$ are taken to be null at $i=1$.

Each $C_i(y,x)$ represents the expected purchasing, holding, and shortage costs incurred for part i given a set of order-up-to levels y_i. We have written $C_i(y,x)$ as the sum of eight terms. The first two terms express the expected purchasing cost, where the first term is the expected purchasing cost if the realized uptime of the machine is such that the available supply of the ith part is between the values of 0 and $y_i - x_i$ and the second term is the expected purchasing cost if the realized uptime of the machine is such that the available supply of the ith part is the desired value $y_i - x_i$. There is no purchasing cost if the available supply of the ith part is not greater than zero. The next three terms represent the expected shortage costs. The first of these terms is the expected shortage cost if the available supply is between 0 and $y_i - x_i$, the second term is the expected shortage cost if the available supply is $y_i - x_i$, and the third is the expected shortage cost if the available supply is 0. Similarly, the last three terms represent the expected holding costs, where the first of these terms is the expected holding cost if the available supply of the ith part is between 0 and $y_i - x_i$, the second term is the expected holding cost if the available supply is $y_i - x_i$, and the third is the expected holding cost if the available supply is 0.

12.4 Properties of the Objective Function

To obtain the optimal order quantities, we wish to show that the total cost function is convex with respect to the order quantities. If this is so, we can find minimizing order quantities by finding where the partial derivative of the total cost function is zero.

12.4.1 First Order Optimality Condition

We begin by finding the partial derivative of the total cost function with respect to y_N. Using Leibnitz's rule, we obtain

$$\frac{\partial}{\partial y_N}C(y,x) = (c_N - p_N\bar{G}_N(y_N) + h_N\bar{G}_N(y_N))\bar{F}\left(\sum_{j=1}^{N} y_j - x_j; T_N\right). \tag{12.6}$$

When written as the product of two terms as we have done, this derivative has a nice interpretation. The first term is the derivative of the cost function for the classical newsboy problem. This term is multiplied by the probability that we can complete our production plan in the time available.

Because of this structure, the first order optimality condition is reduced to $G_N(y_N) = (p_N - c_N)/(p_N + h_N)$, the solution to the classical newsboy problem. As before, it is easy to show that if this implies $y_N < x_N$, then the optimal y_N is x_N. The optimal y_N should not be dependent on the other y_i, because once we have produced parts $1,\ldots,N-1$, all we can do is try to minimize the costs for part N. The optimal y_N should not be dependent on the machine's reliability, because the best thing to do is attempt to achieve the optimal order-up-to quantity exactly.

We now turn to the more difficult task of taking the partial derivative of the total cost function $C(y,x)$ with respect to y_i for $i < N$. After simplification, the result is

$$\begin{aligned}\frac{\partial}{\partial y_i}C(y,x) = {} & \bar{F}\left(\sum_{j=1}^{i} y_j - x_j; T_i\right)(c_i - p_i\bar{G}_i(y_i) + h_iG_i(y_i)) \\ & + \sum_{k=i+1}^{N}(p_k - c_k)\left[F\left(\sum_{j=1}^{k} y_j - x_j; T_k\right) - F\left(\sum_{j=1}^{k-1} y_j - x_j; T_k\right)\right] \\ & - \sum_{k=i+1}^{N}(p_k + h_k)\int_{x_k}^{y_k} G_k(u) f\left(\sum_{j=1}^{k-1} y_j - x_j + u - x_k; T_k\right)\mathrm{d}u, \end{aligned} \tag{12.7}$$

where the summations from $i+1$ to N are taken to be null at $i = N$. This expression is easier to interpret if we rewrite it as

$$\begin{aligned}\frac{\partial}{\partial y_i}C(y,x) = {} & \bar{F}\left(\sum_{j=1}^{i} y_j - x_j; T_i\right)(c_i - p_i\bar{G}_i(y_i) + h_iG_i(y_i)) \\ & + \sum_{k=i+1}^{N}(p_k - c_k)\left[F\left(\sum_{j=1}^{k} y_j - x_j; T_k\right) - F\left(\sum_{j=1}^{k-1} y_j - x_j; T_k\right)\right]\end{aligned}$$

$$+\sum_{k=i+1}^{N}(p_k+h_k)\int_{x_k}^{y_k}\bar{G}_k(u)f\left(\sum_{j=1}^{k-1}y_j-x_j+u-x_k;T_k\right)\mathrm{d}u$$

$$-\sum_{(k)=i+1}^{N}(p_k+h_k)\int_{x_k}^{y_k}f\left(\sum_{j=1}^{k-1}y_j-x_j+u-x_k;T_k\right)\mathrm{d}u, \quad (12.8)$$

and then simplify to obtain

$$\frac{\partial}{\partial y_i}C(y,x)=\bar{F}\left(\sum_{j=1}^{i}y_j-x_j;T_i\right)(c_i-p_i\bar{G}_i(y_i)+h_iG_i(y_i))$$
$$-\sum_{k=i+1}^{N}(c_k+h_k)\left[F\left(\sum_{j=1}^{k}y_j-x_j;T_k\right)-F\left(\sum_{j=1}^{k-1}y_j-x_j;T_k\right)\right]$$
$$+\sum_{k=i+1}^{N}(p_k+h_k)\int_{x_k}^{y_k}\bar{G}_k(u)f\left(\sum_{j=1}^{k-1}y_j-x_j+u-x_k;T_k\right)\mathrm{d}u. \quad (12.9)$$

The first term is analogous to $\partial C(y,x)/\partial y_N$ discussed above. The second two terms give the impact of the choice of y_i on the parts $k=i+1,\ldots,N$. The first of these terms represents the marginal cost of machine time. The expression in square brackets is the probability that machine output is insufficient to produce up to y_k but sufficient to start production of part k. As this probability increases, total cost decreases at rate c_k+h_k, assuming that the units built are not sold. The final term is the marginal cost of lost sales. The integral represents the expected sales given that machine output is greater than zero but less than y_k. As this increases, shortage costs are accrued at a rate p_k and holding costs, which have already been charged in the second term, are avoided at a rate h_k.

It can be seen from the first order condition that as T tends to infinity, the optimal y_i each approach their "newsboy point" y_i^N, that is, the point where $G(y_i)=(p_i-c_i)/(p_i+h_i)$. However, we can prove a stronger result, as stated by the following:

Theorem 1. *The optimal y_i are never greater than y_i^N, their respective newsboy points.*

Proof. We have already shown that the optimal y_N is y_N^N, the newsboy point for part N. Suppose that we have shown that the optimal y_k are not greater than y_k^N for $k=i+1,\ldots,N$. We will now show that the optimal y_i is also less than or equal to y_i^N. We first require the following.

Lemma 1.

$$\frac{\partial}{\partial y_i}C(y,x)\geq\bar{F}\left(\sum_{j=1}^{i}y_j-x_j;T_i\right)(c_i-p_i\bar{G}_i(y_i)+h_iG_i(y_i))$$

Proof.

$$\frac{\partial}{\partial y_i}C(y,x) = \bar{F}\left(\sum_{j=1}^{i} y_j - x_j; T_i\right)(c_i - p_i\bar{G}_i(y_i) + h_iG_i(y_i))$$
$$+ \sum_{k=i+1}^{N}(p_k - c_k)\left[F\left(\sum_{j=1}^{k} y_j - x_j; T_k\right) - F\left(\sum_{j=1}^{k-1} y_j - x_j; T_k\right)\right]$$
$$- \sum_{k=i+1}^{N}(p_k + h_k)\int_{x_k}^{y_k} G_k(u) f\left(\sum_{j=1}^{k-1} y_j - x_j + u - x_k; T_k\right)\mathrm{d}u$$

(from equation (12.7))

$$\geq \bar{F}\left(\sum_{j=1}^{i} y_j - x_j; T_i\right)(c_i - p_i\bar{G}_i(y_i) + h_iG_i(y_i))$$
$$+ \sum_{k=i+1}^{N}(p_k - c_k)\left[F\left(\sum_{j=1}^{k} y_j - x_j; T_k\right) - F\left(\sum_{j=1}^{k-1} y_j - x_j; T_k\right)\right]$$
$$- \sum_{k=i+1}^{N}(p_k + h_k)G_k(y_k)\int_{x_k}^{y_k} f\left(\sum_{j=1}^{k-1} y_j - x_j + u - x_k; T_k\right)\mathrm{d}u$$

(because $G_k(\cdot)$ is nondecreasing)

$$= \bar{F}\left(\sum_{j=1}^{i} y_j - x_j; T_i\right)(c_i - p_i\bar{G}_i(y_i) + h_iG_i(y_i))$$
$$- \sum_{k=i+1}^{N}(c_k - p_k\bar{G}_k(y_k) + h_kG_k(y_k))$$
$$\times\left[F\left(\sum_{j=1}^{k} y_j - x_j; T_k\right) - F\left(\sum_{j=1}^{k-1} y_j - x_j; T_k\right)\right]$$
$$\geq \bar{F}\left(\sum_{j=1}^{i} y_j - x_j; T_i\right)(c_i - p_i\bar{G}_i(y_i) + h_iG_i(y_i)). \tag{12.10}$$

(because $G_k(y_k) \leq (p_k - c_k)/(p_k + h_k)$ for $k = i+1,\ldots,N$, by the induction hypothesis). □

Using this result, it immediately follows that for any $y_i > y_i^N, \partial C(y,x)/\partial y_i$ is positive. Therefore, if $x_i < y_i^N$, the optimal y_i lies between x_i and y_i^N. If $x_i \geq y_i^N$, then it is optimal not to produce (the optimal y_i equals x_i).

12.4.2 Convexity of the Total Cost Function

In this section, we prove the following:

Theorem 2. *If* $x_k \leq y_k \leq y_k^N$ *for* $k = i, i+1, \ldots, N$, *then* $\partial^{\mathrm{y}} C(y,x)/\partial y_i^{\mathrm{y}}$ *is nonnegative.*

Proof. We once again use Leibnitz's rule to take the second partial derivative with respect to y_i to obtain

$$\begin{aligned}\frac{\partial^2}{\partial y_i^2} C(y,x) = & \left[\bar{F}\left(\sum_{j=1}^{i} y_j - x_j; T_i\right)(p_i g_i(y_i) + h_i g_i(y_i))\right] \\ & + \left[-f\left(\sum_{j=1}^{i} y_j - x_j; T_i\right)(c_i - p_i \bar{G}_i(y_i) + h_i G_i(y_i))\right] \\ & + \left[\sum_{k=i+1}^{N}\left\{(p_k - c_k)\left(f\left(\sum_{j=1}^{k} y_j - x_j; T_k\right) - f\left(\sum_{j=1}^{k-1} y_j - x_j; T_k\right)\right)\right.\right. \\ & \left.\left. -(p_k + h_k)\int_{x_k}^{y_k} G_k(u)\frac{\partial}{\partial y_i}\left\{f\left(\sum_{j=1}^{k-1} y_j - x_j + u - x_k; T_k\right)\right\}\mathrm{d}u\right\}\right].\end{aligned} \tag{12.11}$$

We now show that this second partial derivative is nonnegative. We have written the second partial derivative as the sum of three (square bracketed) terms. The first term can be seen to be nonnegative by inspection. The second square bracketed term is nonnegative if $-c_i + p_i\bar{G}_i(y_i) - h_i G_i(y_i)$ is nonnegative, which is true if $G_i(y_i) \leq (p_i - c_i)/(p_i + h_i)$, which is always true for $y_i < y_i^N$. Showing that the third bracketed term is nonnegative is slightly more difficult. We note that for each k,

$$\begin{aligned}&(p_k - c_k)\left(f\left(\sum_{j=1}^{k} y_j - x_j; T_k\right) - f\left(\sum_{j=1}^{k-1} y_j - x_j; T_k\right)\right) \\ &\quad -(p_k + h_k)\int_{x_k}^{y_k} G_k(u)\frac{\partial}{\partial y_i}\left\{f\left(\sum_{j=1}^{k-1} y_j - x_j + u - x_k; T_k\right)\right\}\mathrm{d}u \\ &\geq (p_k - c_k)\left(f\left(\sum_{j=1}^{k} y_j - x_j; T_k\right) - f\left(\sum_{j=1}^{k-1} y_j - x_j; T_k\right)\right) \\ &\quad -(p_k + h_k)G_k(y_k)\int_{x_k}^{y_k} \frac{\partial}{\partial y_i}\left\{f\left(\sum_{j=1}^{k-1} y_j - x_j + u - x_k; T_k\right)\right\}\mathrm{d}u\end{aligned}$$

(because $G_k(\cdot)$ is nondecreasing)

$$\geq (p_k - c_k)\left(f\left(\sum_{j=1}^{k} y_j - x_j; T_k\right) - f\left(\sum_{j=1}^{k-1} y_j - x_j; T_k\right)\right)$$
$$-(p_k + h_k)\frac{(p_k - c_k)}{(p_k + h_k)}\int_{x_k}^{y_k} \frac{\partial}{\partial y_i}\left\{f\left(\sum_{j=1}^{k-1} y_j - x_j + u - x_k; T_k\right)\right\}\mathrm{d}u$$

(because $G_k(y_k) \leq (p_k - c_k)/(p_k + h_k)$)

$$= (p_k - c_k)\left(f\left(\sum_{j=1}^{k} y_j - x_j; T_k\right) - f\left(\sum_{j=1}^{k-1} y_j - x_j; T_k\right)\right)$$
$$-(p_k - c_k)\left(f\left(\sum_{j=1}^{k} y_j - x_j; T_k\right) - f\left(\sum_{j=1}^{k-1} y_j - x_j; T_k\right)\right) = 0. \quad (12.12)$$

□

Given the other $y_j, j \neq i$, this result allows us to find the optimal y_i by determining if $\exists\ y_i \in [x_i, y_i^N]$ such that $\partial C(y,x)/\partial y_i = 0$. If such a y_i exists then it is optimal, otherwise, the optimal policy is not to order. Since $\partial C(y,x)/\partial y_i$ is a nondecreasing function of y_i over the range $[x_i, y_i^N]$ when $y_k \leq y_k^N$ for $k = i+1, \ldots, N$, the optimal y_i can be found by simple binary search.

Given the above results, after we have found y_N we can find the other y_i by solving the above problem as a $N-1$ dimensional unconstrained minimization problem on the interval $x_i \leq y_i \leq y_i^N, i = 1, \ldots, N-1$. For a discussion of general algorithms to solve such problems, see Bazaraa et al. (1993). Below we present a solution algorithm that exploits the special structure of the model.

12.5 Solution Algorithm

The difficulty in finding the optimal production quantities is that the first order condition tells us that $N-1$ of the y_i are mutually dependent. We now describe a solution procedure that exploits the special structure of these dependencies. In particular, consider the difference

$$\hat{C}_{i+1} = \frac{\partial C(y,x)}{\partial y_{i+1}} - \frac{\partial C(y,x)}{\partial y_i}$$
$$= \bar{F}\left(\sum_{j=1}^{i+1} y_j - x_j; T_{i+1}\right)(c_{i+1} - p_{i+1}\bar{G}_{i+1}(y_{i+1}) + h_{i+1}G_{i+1}(y_{i+1}))$$

$$-\bar{F}\left(\sum_{j=1}^{i} y_j - x_j; T_i\right)(c_i - p_i\bar{G}_i(y_i) + h_i G_i(y_i))$$
$$-(p_{i+1} - c_{i+1})\left[F\left(\sum_{j=1}^{i+1} y_j - x_j; T_{i+1}\right) - F\left(\sum_{j=1}^{i} y_j - x_j; T_{i+1}\right)\right]$$
$$+(p_{i+1} + h_{i+1})\int_{x_{i+1}}^{y_{i+1}} G_{i+1}(u) f\left(\sum_{j=1}^{i} y_j - x_j + u - x_{i+1}; T_{i+1}\right) \mathrm{d}u. \tag{12.13}$$

Note that if y_i is optimal, $\partial C(y,x)/\partial y_i$ is zero, so that $\hat{C}_{i+1} = C(y,x)/y_{i+1}$. The reason that this is significant is because $\hat{C}_{i+1}$ is a function only of $y_1,\ldots,y_i$. Therefore if the optimal y_1 is known then $\hat{C}_2$ can be used to find the optimal y_2, and then $\hat{C}_3$ can be used to find the optimal y_3, and so forth.

Since the optimal y_1 is not known, we must use a search technique to find it. We now prove three important properties that will be helpful in this regard.

Let the production quantities that result from the above procedure be denoted by $\hat{y}_i$. We first show that $\hat{y}_N = y_N^N$ iff $\partial C(y,x)/\partial y_1 = 0$. Observe that $\hat{C}_N$ is exactly equal to $\partial C(y,x)/\partial y_N - \partial C(y,x)/\partial y_{N-1}$, and thus $\hat{y}_N = y_N^N$ iff $\partial C(y,x)/\partial y_{N-1} = 0$. Further, for any i, $\hat{C}_{i+1} = C(y,x)/y_{i+1}$ iff $\partial C(y,x)/\partial y_i = 0$. Therefore, $\hat{y}_N = y_N^N$ iff $\partial C(y,x)/\partial y_1 = 0$.

The second property is that if the guess for the optimal value of y_1 is too large, $\hat{y}_N > y_N^N$. We have shown above that if $x_k \le y_k \le y_i^N$ for $k = i, i+1,\ldots,N$, then $\partial^2 C(y,x)/\partial y_i^2 \ge 0$. Accordingly, if the guess for the optimal value of y_1 is too large, $\partial C(y,x)/\partial y_1 > 0$, so that in order for $\hat{C}_2 = 0, \hat{y}_2$ must be chosen such that $\partial C(y,x)/\partial y_2 > 0$, so that $\hat{y}_2$ will be greater than the optimal y_2. Repeating this argument, we see that each $\hat{y}_i$ will be greater than the optimal y_i, and thus $\hat{y}_N > y_N^N$. By analogous reasoning, we can conclude that if the guess for the optimal value of y_1 is too small, $\hat{y}_N < y_N^N$.

The third and final property that we wish to show is that $\hat{C}_{i+1}$ is an increasing function of y_{i+1}. This property is particularly important, as it allows us to find $\hat{y}_{i+1}$ by simple binary search. To prove this, we take the partial derivative of $\hat{C}_{i+1}$ with respect to y_{i+1} and simplify to obtain

$$\frac{\partial}{\partial y_{i+1}}\hat{C}_{i+1} = \bar{F}\left(\sum_{j=1}^{i+1} y_j - x_j; T_{i+1}\right)(p_{i+1} g_{i+1}(y_{i+1}) + h_{i+1} g_{i+1}(y_{i+1})), \tag{12.14}$$

which is clearly nonnegative since each term is nonnegative, and thus the result is proven.

Using these properties, we are now ready to state the following:

Algorithm:

1. Preprocessing. Compute the y_i^N. If any $x_i \geq y_i^N$, then the optimal $\hat{y}_i = x_i$ and it is optimal not to produce this part. Remove all such parts from the list of parts to be produced over the horizon.
2. Initialization. Set $\hat{y}_1 = y_1^N$. Set $U = y_1^N$ and $L = x_1$.
3. Main loop. For each $i = 2,\ldots,N$, find the $\hat{y}_i$ such that $\hat{C}_i = 0$. If any $\hat{y}_i > y_i^N$, then $\hat{y}_1$ is too large. Set $U = \hat{y}_1$, $\hat{y}_1 = (U+L)/2$, and repeat Step 3.
4. Optimality test. If $|\hat{\mathrm{y}}_\mathrm{N} - \mathrm{y}_\mathrm{N}^\mathrm{N}| < \varepsilon$, then the $\hat{y}_i$ are optimal. Stop.
5. Adjustment step. If $\hat{y}_N > y_N^N$, then $\hat{y}_1$ is too large. Set $U = \hat{y}_1$, $\hat{y}_1 = (U+L)/2$, and go to Step 3. If $\hat{y}_N < y_N^N$, then $\hat{y}_1$ is too small. Set $L = \hat{y}_1$, $\hat{y}_1 = (U+L)/2$, and go to Step 3.

The algorithm essentially performs a binary search on the guess for the optimal y_1 by maintaining an upper and lower bound (U and L) on the optimal value. The algorithm terminates when the current value of $\hat{y}_N$ is within some small positive ε of y_N^N.

Because the properties that we have proven above are valid only if $x_i \leq y_i \leq y_i^N$ for $i = 1,\ldots,N$, we must take care to ensure that this remains true throughout the algorithm. We perform the test in Step 2 to ensure that we do not proceed if any $y_i > y_i^N$. We set $L = x_1$ so that $\hat{y}_1 \geq x_1$. Lastly, in a preprocessing step we remove a part i from consideration if $x_i > y_i^N$. We can do this because, for any such part, the optimal $\hat{y}_i$ is x_i, and it is thus optimal not to produce that part. Since the part would not be produced, it has no effect on the other parts.

12.6 Numerical Results

In this section, we present numerical results from an implementation of the solution algorithm described in the previous section.

For simplicity we describe a two part ($N = 2$) system, with identical parameters for the two parts. The base case parameters used are summarized in Table 12.2. We assume that the demand distribution g is normally distributed with mean equal to 100 and standard deviation equal to 10. We also assume that f, the distribution of output of the machine over a horizon of length T, has a mean of T a standard deviation of $0.1T$, which equates to a coefficient of variation of 0.1.

In this two part example, we know that y_2^* will equal its newsvendor point, which in this case is equal to 97.47. To find y_1^*, we developed a simple spreadsheet model that computes $\hat{C}_i$. The solution procedure then simply searches over values of y_1 until $\hat{C}_2 = 0$. The solution procedure converges to a value equal to the optimal solution within four digits of precision in just 13 iterations. In this case, the value is $y_1^* = 85.31$.

Table 12.2 Base case parameters for experiments

Parameter	Value
T	160
x_i	0
c_i	2
h_i	1
p_i	4
s_i	0

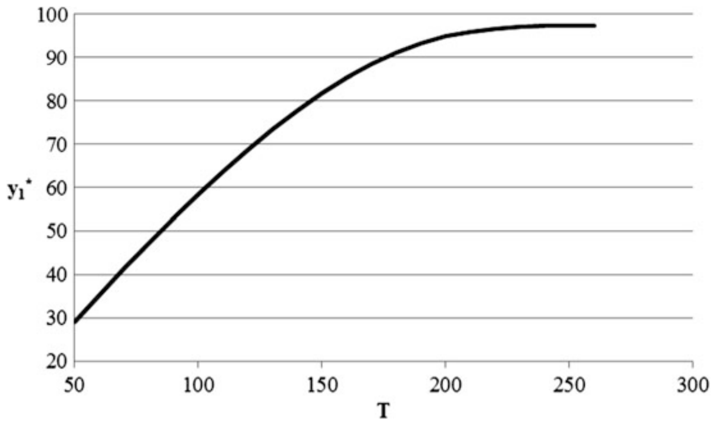


Fig. 12.1 Optimal production quantity of part 1 as a function of T

Next, we wish to show the effect of the time constraint on the optimal value for y_1. In Fig. 12.1 we show the results of varying the value of T, which also impacts our production distribution f. As expected, the value of T has a major effect on the production schedule, until T becomes sufficiently large, at which point y_1^* approaches its newsvendor point.

Finally, in Fig. 12.2 we show the effect of varying the coefficient of variation in the production schedule. This was set to 0.1 in the base case. We see that increased variability has the effect of causing greater levels of planned production as a hedge against this uncertainty.

12.7 Extensions to the Model

In this section, we present three simple extensions to the model: (1) a method for rescheduling within the planning horizon, (2) an extension to evaluate whether or not to purchase the option to run overtime within the planning horizon, and (3) an extension that permits the modeling of a machine that operates at a different speed depending on the part being produced.

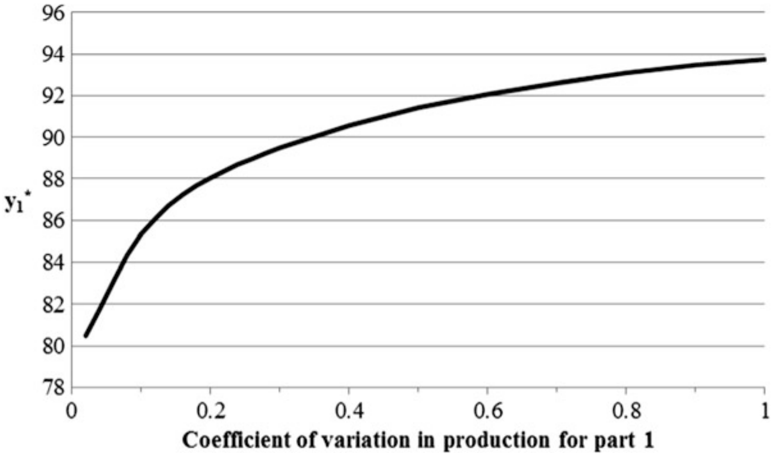


Fig. 12.2 Optimal production quantity of part 1 as a function of the variability in production

12.7.1 Dynamic Rescheduling

In the development above, we have discussed how to determine a set of production quantities to minimize expected total cost. We have assumed up to this point that this plan, once established, is fixed. That is, the plan is implemented by producing the predetermined optimal quantity of each part, and then switching production to the next part. Of course, as this plan is implemented, the reliability of the machine may be much higher or much lower than expected. As a result, if we were given the opportunity to do so, we might adjust the production plan based on the actual realized reliability of the machine.

Suppose we are now permitted to modify our choice of production quantity for the current part. We propose a simple method that allows someone on the shop floor to determine when to stop production of the current part based on the part's inventory level. We denote this production level as the *critical inventory level*. The method described below could produce a chart with two axes: the horizontal axis will be time, and the vertical axis will be the critical inventory level. In this sense, this method could produce a visual aide for a production manager on the factory floor.

Suppose for a particular future point in time t_1 we would like to determine the amount of inventory at or above which it is optimal to stop producing the current part and switch to the next part. We denote this production level as the *critical inventory level*, and determine it as follows. The first step is to update the horizon length in the equations above by replacing T with $T - t_1$. This reflects the amount of time that would be remaining for production at time t_1. Next, search over values for x_1, at each iteration finding the optimal y_i, until we identify the *lowest* x_1 such that at the optimum, $y_1 = x_1$. This is the critical inventory level, since this is the inventory level at which is it optimal not to produce.

We now prove the existence of such a critical inventory level. Recall that we are only interested in the *lowest* x_1 such that at the optimum $y_1 - x_1 = 0$, so the only question we must answer is whether or not such an x_1 exists. But this is clearly so, since if we set $x_1 = G_1{}^{-1}((p_1 - c_1)/(p_1 + h_1))$, we know $y_1 \leq G_1{}^{-1}((p_1 - c_1)/(p_1 + h_1))$ and since we must constrain y_1 to be at least $x_1, y_1 = x_1$.

We can vary the value of t_1 and find the critical inventory levels at each point over the planning horizon. The optimal dynamic operating policy is therefore implemented on the shop floor by producing until the inventory level crosses this curve. Once this happens and production is switched to the next part, the model should be solved again to find the critical inventory level as a function of time for the next part.

12.7.2 Including Options to Run Overtime

In the development above, we purposely omitted any discussion of how to make optimal overtime decisions. Suppose now that there are $m = 1, \ldots, N_{\text{OT}}$ opportunities over the horizon to run overtime, and for simplicity assume that they are each of duration OT at cost w_{m}. Without loss of generality, we will assume that the opportunities are indexed in the order of increasing cost.

In the development above, we computed optimal production quantities ignoring overtime opportunities. This is equivalent to assuming that we choose not to run overtime, and the resulting expected cost is the expected cost of this strategy. Suppose instead that we decide that we are going to run overtime once. To evaluate the expected cost of this strategy we simply replace T by $T + \text{OT}$ and find the optimal production quantities to compute the minimum expected cost, and then add w_{m}. Note that it does not matter where within the planning horizon that we run overtime, since all overtime opportunities occur before the demand point. As a result, we can find the optimal policy by simply running the solution algorithm above $N_{\text{OT}} + 1$ times, with T taking on the values $T, T + \text{OT}, T + 2\text{OT}, \ldots, T + N_{\text{OT}}$ OT, and choosing the strategy with lowest expected cost. Intuitively, increasing the length of the horizon will have a non-increasing benefit. If this is true, which we leave as a conjecture, the evaluation of policies can be stopped when the total cost increases from the previous iteration.

12.7.3 Extension to Different Machine Speeds

For notational convenience, up to this point we have assumed that the machine operates at the same speed when producing different parts. If the speeds are different, then the requirements on the machine need to be expressed in common units, such as time, instead of parts. This can be accommodated easily, replacing all expressions such as $F\left(\sum_{j=1}^{i} y_j - x_j; T_i\right)$ with $F\left(\sum_{j=1}^{i} \frac{y_j - x_j}{P_j}; T_i\right)$, where P_j is the speed at which the machine produces part j when it is working. Our solution procedure for finding the optimal y_i is unchanged by this modification.

12.8 Conclusion and Future Research

In this chapter, we have extended the basic newsvendor model to an unreliable machine that must produce multiple parts in a given period of time. We have seen that with an infinite production horizon, the problem simply decomposes into a single item newsvendor problem for each part. However, under a time constraint, the optimal production quantities are reduced from those in the infinite horizon case. We showed that the optimal production quantities are mutually dependent on one another, but have discovered a special structure in these relationships, and also proven other important properties of the objective function and decision variables. These results allowed us to construct a simple algorithm that performs a binary search for the optimal value of the production quantity for the first part. In each iteration, we exploit the special structure of the problem to easily determine the production quantities for the other parts, and easily test if the overall solution is optimal.

Our formulation has assumed that the production sequence is fixed. This could be the result of sequence-dependent setup costs or setup times, or a function of the timing of arrivals of materials from upstream suppliers. A future research topic could include relax this assumption and allow the decision maker to change the sequence, possibly with penalty costs associated with changes. We have also assumed that demand is satisfied for all parts at the end of the horizon. Another future research topic could be an extension to multiple time periods, or allowing different parts to have demand pull from inventory at different points in time.

Acknowledgments This chapter is based on the author's unpublished Ph.D. dissertation. The author would like to thank Steve Graves for his support of this research, for many helpful suggestions, and comments on earlier drafts of this manuscript. Don Rosenfield and Vien Nguyen provided valuable input on earlier drafts. Dr. Dave Vander Veen and Dr. Bill Jordan of General Motors provided the author with the opportunity to study the real-world problems that were the inspiration for this work. The author gratefully acknowledges the support and resources made available to him through the Leaders for Manufacturing Program (currently the Leaders for Global Operations Program), a partnership between MIT and global manufacturing companies. The author also appreciates the feedback and suggestions from two anonymous referees on an earlier version of this chapter.

References

Al-Salamah, M., & Abudari, A. (2012). Production lot sizing & process targeting under process deterioration & machine breakdown conditions. *Modeling and Simulation in Engineering*, Hindawi Publishing Corporation, 7 pages doi:10.1155/2012/393495

Bazaraa, M. S., Sherali, H. D., & Shetty, C. M. (1993). *Nonlinear programming: Theory and algorithms* (2nd ed.). New York: Wiley.

Beyer, W. H. (1987). *CRC handbook of mathematical sciences* (6th ed.). Boca Raton: CRC Press.

Birge, J., Frenk, J. B. G., Mittenthal, J., & Rinnooy Kan, A. H. G. (1990). Single-machine scheduling subject to stochastic breakdowns. *Naval Research Logistics, 37*, 661–677.

Birge, J. R., & Glazebrook, K. D. (1988). Assessing the effects of machine breakdowns in stochastic scheduling. *Operations Research Letters, 7*(6), 267–271.

Bitran, G. R., Haas, E. A., & Matsuo, H. (1986). Production planning of style goods with high setup costs and forecast revisions. *Operations Research, 34*(2), 226–236.

Ciarallo, F. W., Akella, R., & Morton, T. E. (1994). A periodic review, production planning model with uncertain capacity and uncertain demand. *Management Science, 40*(3), 320–332.

Dada, M., Petruzzi, N. C., & Schwarz, L. B. (2007). A newsvendor's procurement problem when suppliers are unreliable. *Manufacturing & Service Operations Management, 9*(1), 9–32.

Diedrich, F., Jansen, K., Schwarz, U. M., & Trystram, D. (2009) A survey on approximation algorithms for scheduling with machine unavailability. In Algorithmics of large and complex networks: design, analysis, and simulation (pp. 50–64). Springer, Berlin.

El-Ferik, S. (2008). Economic production lot-sizing for an unreliable machine under imperfect age-based maintenance policy. *European Journal of Operational Research, 186*(1), 150–163.

Epstein, L., Levin, A., Marchetti-Spaccamela, A., Megow, N., Mestre, J., Skutella, M., et al. (2010). *Universal sequencing on an unreliable machine. Proceedings of the 14th international conference on integer programming and combinatorial optimization* (IPCO XIV). Springer, Berlin.

Evans, R. V. (1967). Inventory control of a multiproduct system with a limited production resource. *Naval Research Logistics Quarterly, 14*(2), 173–184.

Giri, B. C., Yun, W. Y., & Dohi, T. (2005). Optimal lot sizing in an unreliable two-stage serial production-inventory system. *International Transactions in Operational Research, 12*(1), 63–82.

Giri, B. C., & Dohi, T. (2004). Optimal lot sizing for an unreliable production system based on net present value approach. *International Journal of Production Economics, 92*(2), 157–167.

Gittins, J. C. (1979). Bandit processes and dynamic allocation indices. *Journal of the Royal Statistical Society, Series B, 41*, 148–177.

Glazebrook, K. D. (1984). Scheduling stochastic jobs on a single machine subject to breakdowns. *Naval Research Logistics Quarterly, 31*, 251–264.

Groenevelt, H., Pintelon, L., & Seidmann, A. (1992a). Production batching with machine breakdowns and safety stocks. *Operations Research, 40*(5), 959–971.

Groenevelt, H., Pintelon, L., & Seidmann, A. (1992b). Production lot sizing with machine breakdowns. *Management Science, 38*(1), 104–123.

Halim, K. A., Giri, B. C., & Chaudhuri, K. S. (2010). Lot sizing in an unreliable manufacturing system with fuzzy demand and repair time. *International Journal of Industrial and Systems Engineering, 5*(4), 485–500.

Huggins, E. L., & Olsen T. L. (2010). Inventory control with generalized expediting. *Operations Research, 58*(5), 1414–1426.

Jain, K., & Silver, E. A. (1995). The single period procurement problem where dedicated supplier capacity can be reserved. *Naval Research Logistics Quarterly, 42*, 915–934.

Kletter, D. B. (1994). *Determining Production Lot Sizes and Safety Stocks for an Automobile Stamping Plant*. S.M. Thesis, MIT.

Kletter, D. B. (1996) *Planning and Control of an Unreliable Machine In a Multi-item Production-Inventory System*. Ph.D. Thesis, MIT.

Lee, C. Y. (2004). Machine scheduling with availability constraints. In J. Y.-T. Leung (Ed.), *Handbook of scheduling*. Boca Raton: CRC Press.

Matsuo, H. (1990). A stochastic sequencing problem for style goods with forecast revisions and hierarchical structure. *Management Science, 36*(3), 332–347.

Pinedo, M., & Rammouz, E. (1988). A note on stochastic scheduling on a single machine subject to breakdown and repair. *Probability in the Engineering and Informational Sciences, 2*, 41–49.

Pinedo, M. (2008). *Scheduling: Theory, Algorithms, and Systems* (3rd ed.). Englewood Cliffs: Prentice-Hall.

Racke, H. (2009). Survey on oblivious routing strategies. In K. Ambos-Spies, B. Lowe, & W. Merkle (Eds.), *Mathematical theory and computational practice, Proceedings of 5th Conference on Computability in Europe (CiE)* (pp. 419–429). Springer, Berlin.

Reiman, M. I., & Wein, L. M. (1998). Dynamic scheduling of a two-class queue with setups. *Operations Research, 46*(4), 532–547.

Rose, J. S. (1992). The newsvendor with known demand and uncertain replenishment: applications to quality control and container fill. *Operations Research Letters, 11*(2), 111–117.

Schmidt, G. (2000). Scheduling with limited machine availability. *European Journal of Operational Research, 121*, pp. 1–15.

Sethi, S. P., & Zhang, Q. (1994). *Hierarchical production and setup scheduling in a single machine system, Chapter 8 in hierarchical decision making in stochastic manufacturing systems*. Boston: Birkhäuser.

Smith, S. A., Chambers, J. C., & Shlifer, E. (1980). Optimal inventories based on job completion rate for repairs requiring multiple items. *Management Science, 26*(8), 849–852.

Yano, C. A., & Lee, H. L. (1995). Lot sizing with random yields: a review. *Operations Research, 43*(2), 311–334.

Chapter 13
Analysis of the Single-Period Problem under Carbon Emissions Policies

Jingpu Song and Mingming Leng

Abstract We investigate the classical single-period (newsvendor) problem under carbon emissions policies including the mandatory carbon emissions capacity, the carbon emissions tax, and the cap-and-trade system. Specifically, under each policy, we find a firm's optimal production quantity and corresponding expected profit, and draw analytic managerial insights. We show that, in order to reduce carbon emissions by a certain percentage, the tax rate imposed on the high-margin firm should be less than that on the low-margin firm for the high-profit perishable products, whereas the high-margin firm should absorb a high tax than the low-margin firm for the low-profit products. Under the cap-and-trade policy, the emissions capacity should be set to a level such that the marginal profit of the firm is less than the carbon credit purchasing price. We also derive the specific (closed-form) conditions under which, as a result of implementing the cap-and-trade policy, the firm's expected profit is increased and carbon emissions are reduced.

Keywords Cap-and-trade • Carbon emissions • Carbon tax • Single-period model

13.1 Introduction

The past three decades have clearly witnessed an increasingly serious impact of carbon dioxide on the environment. Carbon dioxide has been regarded as the main

J. Song
International Business School, Shanghai Institute of Foreign Trade,
1900 Wenxiang Road, Songjiang, Shanghai, China
e-mail: jingpusong@gmail.com

M. Leng (✉)
Department of Computing and Decision Sciences, Faculty of Business, Lingnan University,
8 Castle Peak Road, Tuen Mun, Hong Kong
e-mail: mmleng@ln.edu.hk

T.-M. Choi (ed.), *Handbook of Newsvendor Problems*, International Series in Operations Research & Management Science 176, DOI 10.1007/978-1-4614-3600-3_13,

pollutant that is warming the Earth. It is a greenhouse gas that is emitted through transport, land clearance, and the production and consumption of food, fuels, manufactured goods, materials, wood, roads, buildings, and services CO2List.org (2006). For the purpose of environmental protection, many governments and organizations have been contributing to carbon emissions reduction with a common goal that carbon emissions should be reduced by at least half by 2050, as reported by, e.g., the International Energy Agency (2008).

In practice, a great number of governments have implemented some policies to control carbon emissions. In Congress of the United States (2008), the Congressional Budget Office (CBO) of the Congress of the United States provided a comprehensive study on the policy options for reducing CO_2 emissions. We find from the CBO's study that there are four *major* carbon emissions policies as follows: (i) a mandatory capacity on the amount of carbon emitted by each firm; (ii) a tax imposed to each firm on the amount of carbon emissions; (iii) a cap-and-trade system implemented to allow the emission trading; and (iv) an investment made by each firm in the carbon offsets to meet its carbon capacity requirement. In Sect. 13.2.2, we shall specify these four major policies, and show that the fourth policy can be per se regarded as a special case of the third Policy and it should be thus necessary, and interesting, to investigate the first, the second, and the third policies.

In this paper, we analyze the impact of the three policies on a profit-oriented firm's production quantity decision. We note that many profit-oriented firms have also observed the importance of the carbon emissions reduction, and responded by developing low-carbon technologies and adopting new and renewable energy resources. Furthermore, the Barloworld Optimus—the logistics arm of the multinational corporation "Barloworld"—reported that, even though over 80% of carbon savings are usually achieved at the product design stage, each firm can reduce carbon emissions by optimizing its operations in production, inventory, and transportation; see, for example, Benjaafar et al. (2010) and BuySmart network (2008). A survey by Accenture.com (2009) indicated that more than 86% supply chain executives have undertaken at least one green initiative in the areas such as recycling, lighting management, and energy-efficient systems. We also learn from Accenture.com (2009) that 10% of companies have actively modeled their supply chain carbon footprints and implemented successful sustainability initiatives.

For our analysis of carbon emissions policies, we focus on the optimal quantity decision of a firm making a perishable item with a short lifespan. The production of the item results in carbon emissions. It is realistic to consider the perishable item for the firm. For example, in the Huber Group (2003), the Huber Group—which provides facility services to commercial, industrial, educational, medical, retail, government, and institutional customers—released a technical information regarding the impact of newspaper printing with the carbon-based ink on the environment. In addition, as reported in Environmental News Energies Correspondent (2009), Carbon Trust, a British governmental organization, suggests that consumers should use real Christmas trees instead of artificial equivalents, because the carbon footprint left by artificial trees is at least ten times greater than real Christmas trees.

However, in today's market, the demand for artificial Christmas trees is still very high; for example, Tesco—the largest British supermarket chain—sold 300,000 artificial Christmas trees in December 2009.

To examine how each carbon emissions policy affects the firm's production quantity decision, we shall involve a corresponding parameter into the classical single-period model, and address the following questions:

1. What are the firm's optimal production quantity decision and corresponding maximum expected profit under each carbon emissions policy?
2. How does the implementation of a policy influence the carbon emissions reduction and the expected profits of the low-margin, the moderate-margin, and the high-margin firms?
3. Does there exist a "win–win" scenario in which the carbon emissions are decreased while the firm's expected profit is not reduced?

Our paper contributes to the literature by analyzing the single-period problem under carbon emissions policies and presenting managerial discussion on the incentive of the firm on the carbon emissions reduction. Even though our discussions on the policies are motivated by the practice of the U.S., our analytic approach and results should be useful to any government who intends to choose a proper policy to reduce carbon emissions. The remainder of this paper is organized as follows: In Sect. 13.2, we briefly review the relevant literature in Sect. 13.2.1, which shows the originality of this paper; and we present our discussion on existing carbon emissions policies in Sect. 13.2.2. In Sect. 13.3, we consider three policies, and for each policy analyze the single-period model to find the corresponding optimal quantity decision. Numerical study with sensitivity analysis are provided in Sect. 13.4. This paper ends with a summary of our results in Sect. 13.5. In addition, a list of major notations used in this paper is given in Table 13.1.

13.2 Preliminaries: Literature Review and Carbon Emissions Policies

In this section, we briefly review major relevant publications and discuss four carbon emissions policies, which are preliminaries to our analysis of the single-period problem under carbon emissions policies.

13.2.1 Brief Literature Review

We now review major publications that are closely related to this paper where we analyze the classical single-period model in the presence of carbon emissions policies. For a detailed description of the classical model, see, e.g., Hadley and Whitin (1963). The single-period model has been widely used to investigate a

Table 13.1 A list of major notations that are used in this paper

Notation	Definition
α	Unit purchasing price of the carbon credits
β	Unit selling price of the carbon credits
c	Unit acquisition cost of the perishable product
c_{o}	Unit overage cost
c_{u}	Unit underage cost
C	Fixed carbon capacity
e	Average carbon emissions per unit of the perishable product
κ	Percentage of the reduction in carbon emissions
Q	Order/production quantity
Q_{c}	Mandatory capacity for carbon emissions
s	Shortage (stockout) cost for each unsatisfied demand
τ	Tax amount paid by the firm for each unit of the perishable product
v	Salvage value per unit of the unsold perishable product
X	Aggregate demand, which is assumed to be a random variable with the probability density function (p.d.f.) $f(x)$ and the cumulative distribution function (c.d.f.) $F(x)$.

variety of problems in the operations management (OM) area. Khouja (1999) proposed a literature review of various single-period problems. In today's OM area, many scholars still extend the classical model to incorporate different objectives and utility functions, address different pricing policies, analyze the value of the demand information, etc.

Starting from the middle of 1990s, the carbon emissions-related issues have been attracting the OM scholars' attention. As a seminal publication, Penkuhn et al. (1997) considered the emission taxes and developed a nonlinear programming model for a production planning problem. Letmathe and Balakrishnan (2005) constructed two analytic models to determine a firm's production quantities under different environmental constraints. Kim et al. (2009) investigated the relationship between transportation costs and CO_2 emissions using the multi-objective optimization method. Cachon (2009) discussed how a reduction in carbon footprints affects supply chain operations and structures.

In recent two years, an increasing number of OM scholars examine some carbon emissions-related issues. For example, Hoen et al. (2010) investigated the effects of two regulation mechanisms on the decision on the transportation mode selection. Benjaafar et al. (2010) discussed how the carbon emissions concerns could be involved into the operational decision-making models with regard to procurement, production, and inventory management. They also provided insights that highlight the impact of operational decisions on the carbon emissions and the importance of the operational models in assessing the benefits of investments in more carbon-efficient technologies. Hua et al. (2010) investigated how firms manage the carbon emissions in their inventory control under the carbon emissions- trading mechanism. They derived the EOQ model, and analytically examined the impact of carbon trade, carbon price, and carbon capacity on order decisions, carbon emissions, and total cost.

Table 13.2 Four major carbon emission policies discussed by the Congressional Budget Office of the Congress of the United States

Policy		Brief Description
Policy 1:	Mandatory carbon emissions capacity	A firm's production quantity Q of the items that emit the carbon cannot exceed the mandatory capacity Q_c.
Policy 2:	Carbon emissions tax	A firm absorbs the tax τ for each unit of the produced item that emits the carbon.
Policy 3:	Cap-and-trade	A firm—with carbon credits prescribed by the policy-maker to allow the firm to make at most Q_c units of the items—can sell its unused credits at the sale price $\$\beta$ per item or buy other firms' extra credits at the purchasing price $\$\alpha$ per item.
Policy 4:	Investment in the carbon offsets	A firm is allowed to invest for the reduction in carbon emissions to meet the requirement of the mandatory capacity Q_c.

In this paper, we consider the classical single-period problem under three carbon emissions policies, which significantly distinguishes our analysis and those by, e.g., Benjaafar et al. (2010) and Hua et al. (2010). Moreover, we quantify the impact of different policies on the emissions reduction and the expected profit of the firm. This further shows the originality of our paper.

13.2.2 Description of Carbon Emissions Policies

We now describe four major carbon emissions policies that are discussed by the Congressional Budget Office of the Congress of the United States (2008). We begin by presenting a summary of these four policies as given in Table 13.2, where Q denotes a firm's production quantity of the items that emit the carbon, and Q_c means the mandatory capacity of the production that results in carbon emissions. Moreover, in Table 13.2, τ represents the tax amount paid by the firm for each unit of the item that emits the carbon; and, β and α denote the firm's unit sale price and unit purchasing price of the carbon credits in the cap-and-trade system, respectively.

Next, we discuss the four policies listed in Table 13.2 to determine which policies shall be later used to analyze the single-period problem. For our single-period problem under Policy 1 ("mandatory carbon emissions capacity"), the firm's optimal decision is subject to the mandatory capacity. That is, the firm needs to determine an optimal production quantity that maximizes its profit under the constraint that the firm's production quantity Q is smaller than or equal to the mandatory capacity Q_c, i.e., $Q \leq Q_c$. Note that, to simplify our analysis and facilitate our managerial discussion, we measure the carbon emissions-related parameters and constraints on the product-unit basis throughout the paper. This is justified as follows: In reality, carbon emissions can be generated from production, transportation, inventory, etc. Letting e denote the average carbon emissions generated by making one unit of product over the single period, we find that, when the firm has to adhere to a fixed carbon capacity C, he cannot produce more than $Q_c = C/e$ products (that is, $Q \leq Q_c$). This implies that it is reasonable to use Q_c instead of C for our analysis of the single-period problem.

For our problem under Policy 2 (“carbon emissions tax”), there is no carbon emissions constraint; but, the firm absorbs the tax on the amount of carbon emissions. Specifically, denoting by τ the carbon tax charged for one unit of product, we can calculate the firm’s total tax payment as τQ. Under Policy 3 (“cap-and-trade”), the firm has prescribed carbon credits from the policy-maker, which allow the firm to produce at most Q_c units of products. However, the firm can trade extra (unused) carbon credits through a cap-and-trade system to vary its carbon capacity. This means that, in the cap-and-trade system, the firm can buy and sell the “right to emit.”

Under Policy 4 (“investment in the carbon offsets”), the firm can invest in the carbon emissions-reduction projects to offset emissions in excess of the capacity Q_c. We note that the investment under Policy 4 is *per se* the same as the credit purchase in a cap-and-trade system under Policy 3 with $\beta = 0$. That is, if the firm’s unused carbon credits cannot be sold, i.e., $\beta = 0$, then Policy 3 is equivalent to Policy 4 because α can be assumed to be the unit investment cost. Hence, Policy 4 can be regarded as a special case of Policy 3. For generality, we do not analyze our single-period problem under Policy 4 in this paper.

According to the above, we subsequently investigate the impact of Policies 1, 2, and 3 on the optimal decision in the single-period problem.

13.3 Analysis of the Single-Period Problem Under Carbon Emissions Policies

In this section, we analyze the classical single-period inventory model under three carbon emissions policies—i.e., Policies 1, 2, and 3 in Table 13.2. Our analytic results are also compared to investigate the impact of the three policies on the reduction in carbon emissions and the firm’s expected profit. Next, we start with the firm’s single-period inventory problem under Policy 1.

13.3.1 The Single-Period Problem Under Policy 1 (Mandatory Carbon Emissions Capacity)

For our analysis of the classical single-period problem, we let X denote the aggregate demand, which is assumed to be a random variable with the probability density function (p.d.f.) $f(x)$ and the cumulative distribution function (c.d.f.) $F(x)$.

In addition, p is the selling price per unit of the perishable product; c is the firm's unit acquisition cost; s is the shortage (stockout) cost for each unsatisfied demand; and v is the salvage value per unit of the unsold product. Then, $c_{\mathrm{o}} \equiv c - v$ is the unit overage cost, and $c_{\mathrm{u}} \equiv p + s - c$ represents the unit underage cost. Note that Q denotes the firm's order quantity, as defined in Table 13.1.

Using the above, we write the firm's expected profit function as,

$$J(Q) = (p - v)\int_0^Q xf(x)\,\mathrm{d}x + (p + s - c)\int_Q^\infty Qf(x)\,\mathrm{d}x - s\int_Q^\infty xf(x)\,\mathrm{d}x - (c - v)\int_0^Q Qf(x)\,\mathrm{d}x. \tag{13.1}$$

We learn from our discussion in Sect. 13.2.2 that, in order to find optimal quantity Q^* under Policy 1 ("mandatory carbon emissions capacity"), the firm should maximize its expected profit $J(Q)$ in (13.1) under the constraint that $Q \leq Q_{\mathrm{c}}$, where Q_{c} is the mandatory capacity. That is, the firm's maximization problem under Policy 1 is written as follows: $\max_{\mathrm{Q} \leq \mathrm{Q_c}} J(Q)$.

Theorem 1. *For the single-period problem under Policy 1 (mandatory carbon emissions capacity), the optimal quantity decision is found as* $Q_1^* = \min(Q^*, Q_{\mathrm{c}})$, *where* Q^* *is optimal solution of the classical single-period problem, i.e.,*

$$Q^* = F^{-1}(w), \quad \text{where } w \equiv \frac{c_{\mathrm{u}}}{c_{\mathrm{u}} + c_{\mathrm{o}}} = \frac{p + s - c}{p + s - v}. \tag{13.2}$$

Proof. For a proof of this theorem and the proofs of all subsequent theorems, see 13.5. □

From the above theorem, we note that Policy 1 is effective only when the mandatory capacity Q_{c} does not exceed the Q^*, i.e., $Q_{\mathrm{c}} \leq Q^*$. Otherwise, if $Q_{\mathrm{c}} > Q^*$, then the firm always determines its optimal solution as Q^* for any value of Q_{c}, which means that the firm's optimal solution under Policy 1 is the same as that with not any policy. It thus follows that, in order to *effectively* reduce carbon emissions generated by the firm, the policy-maker needs to set the mandatory capacity as a value lower than the firm's optimal decision under no policy constraint.

Theorem 1 also indicates that we can compute Q_1^* when the c.d.f. $F(x)$ is explicitly given. For simplicity, we hereafter assume that the aggregate demand X for the perishable product is normally distributed with mean μ and standard deviation σ, i.e., $X \sim N(\mu, \sigma)$. We thus have,

$$J(Q^*) = \mu(p - c) - \sigma(c_{\mathrm{u}} + c_{\mathrm{o}})\phi(z^*), \tag{13.3}$$

where $z^* \equiv (Q^* - \mu)/\sigma$, and ϕ is the p.d.f. of the standard normal distribution.

13.3.2 The Single-Period Problem Under Policy 2 (Carbon Emissions Tax)

Under the policy, the firm needs to pay the tax $\$\tau$ for each unit of product, as discussed in Sect. 13.2.2. This means that the firm incurs the per unit cost $\$\tau$ in addition to its acquisition cost c. Thus, we can easily write the firm's corresponding profit function, by replacing c in $J(Q)$ given in (13.1) with $c+\tau$. As a result, the optimal production quantity under Policy 2 is given as,

$$Q_2^* = F^{-1}\left(\frac{p+s-c-\tau}{p+s-v}\right) = F^{-1}(w_2). \tag{13.4}$$

Next, we discuss the effect of the carbon tax τ on the reduction in carbon emissions. More specifically, we need to consider the following question: what should be the value of τ if we desire to reduce the firm's carbon emissions by a certain percentage. Note that, if Policy 2 does not apply, then the firm's optimal quantity decision is Q^*, as given in (13.2); and, if this policy applies, then the optimal decision is Q_2^* as in (13.4). Therefore, the reduction in carbon emissions can be calculated as $\kappa \equiv (Q^* - Q_2^*)/Q^*$.

In addition, we should also consider the impact of the profitability-related attributes of the perishable product on the policy-maker's tax decision. As discussed by Schweitzer and Cachon (2000), in the single-period problem, the perishable product with $w_2 \geq 0.5$ and that with $w_2 < 0.5$—where w is defined as in Theorem 1—are called a high-profit product and a low-profit product, respectively. Noting that the aggregate demand X follows a normal distribution, we find from (13.2) that $Q^* \geq \mu$ for the high-profit products with $w_2 \geq 0.5$, and $Q^* < \mu$ for the low-profit products with $w_2 < 0.5$. Furthermore, it should be interesting to investigate whether or not the firm selling a high-profit product and the firm selling a low-profit product should have the same tax payment if they desire to achieve a same emission-reduction percentage κ.

Theorem 2. *If the firm makes a high-profit perishable product (i.e., $w_2 \geq 0.5$), then the carbon emissions-reduction percentage κ is* decreasing *in c, i.e., $\partial\kappa/\partial c < 0$. But, if the firm makes a low-profit perishable product (i.e., $w_2 < 0.5$), then the carbon emissions-reduction percentage κ is* increasing *in c, i.e., $\partial\kappa/\partial c > 0$.*

As the above theorem indicates, for a high-profit and a low-profit products under Policy 2 with a fixed value of the carbon tax τ, we find that, *ceteris paribus*, the carbon emissions-reduction percentage κ varies in *different* manners as the unit cost is changed. For a high-profit product, the reduction decreases as c increases, whereas, for a low-profit product, the reduction increases as c increases. The result implies an important insight from the perspective of the policy-maker, as given in the following remark.

Remark 1. The policy-maker should consider the attributes of the perishable product and the unit acquisition cost of the firm, in order to achieve the emissions reduction at a certain desired level. Specifically, for a given value of κ, if the perishable product belongs to the high-profit category, then the tax rate τ imposed on the high-margin firm (i.e., its unit acquisition cost c is small) should be less than that on the low-margin firm (i.e., c is high). On the other hand, for the low-profit product, the high-margin firm should absorb a high tax than the low-margin firm.

13.3.3 The Single-Period Problem Under Policy 3 (Cap-and-Trade)

Under the policy, the firm has to buy the carbon credits at the per unit price α if it produces more than the prescribed capacity Q_c. We thus calculate the purchasing cost of carbon credits as $\alpha\,(Q-Q_c)^+$, where,

$$(Q-Q_c)^+ = \max(Q-Q_c,0) = \begin{cases} Q-Q_c, & \text{if } Q \geq Q_c, \\ 0, & \text{otherwise.} \end{cases} \tag{13.5}$$

Note that, if $Q_c \leq Q$, then $\alpha\,(Q-Q_c)^+ = 0$, which implies that the firm makes no payment if it does not need any extra carbon credits. However, the firm may benefit from emitting less than the capacity Q_c by selling its unused carbon credits in the trading market. In fact, for the single-period problem where the unused credits should be salvaged, the firm has to sell unused credits and thus obtain the revenue as $\beta\,(Q_c-Q)^+$.

Therefore, the firm's expected profit under the cap-and-trade policy can be written as,

$$J_3\,(Q) = J\,(Q) + \alpha(Q-Q_c)^+ + \beta(Q_c-Q)^+, \tag{13.6}$$

where $J\,(Q)$ is given as in (13.1); and as discussed above, the second and third terms can be regarded as the firm's "penalties" and "rewards" generated by transferring carbon credits under the cap-and-trade policy, respectively. The firm should maximize $J_3\,(Q)$ in (13.6) to find the optimal quantity Q_3^* under Policy 3.

Theorem 3. *When Policy 3 ("cap-and-trade") is implemented, we find the firm's optimal quantity decision Q_3^* as given in Table 13.3, where Q^* is the optimal solution for the classical single-period problem, as given in (13.2); and,*

$$w_\alpha \equiv \frac{c_u-\alpha}{c_u+c_o}, \quad w_\beta \equiv \frac{c_u-\beta}{c_u+c_o}, \quad \gamma \equiv \left.\frac{\mathrm{d}J(Q)}{\mathrm{d}Q}\right|_{Q=Q_c}. \tag{13.7}$$

Note that γ in (13.7) means the firm's marginal *profit at the point that $Q = Q_c$. Moreover, the firm's corresponding expected profit is also calculated as in Table 13.3.* ■

Table 13.3 The firm's optimal quantity decision Q_3^* under Policy 3 ("cap-and-trade"). Note that w_α, w_β, and γ are defined as in (13.7)

Condition	$Q_c < Q^*$	$Q_c \geq Q^*$
$\beta \geq c_u$	$Q_3^* = 0$; $J_3(Q_3^*) > J(Q^*)$	$Q_3^* = 0$; $J_3(Q_3^*) > J(Q^*)$
$c_u > \beta > \gamma$	$Q_3^* = F^{-1}(w_\beta) < Q_c$	$Q_3^* = F^{-1}(w_\beta) \leq Q_c$ $J_3(Q_3^*) > J(Q^*)$
$\beta \leq \gamma \leq \alpha$	$Q_3^* = Q_c$; $J_3(Q_3^*) < J(Q^*)$	
$\alpha < \gamma$	$Q_3^* = F^{-1}(w_\alpha) > Q_c$; $J_3(Q_3^*) < J(Q^*)$	

We learn from Theorem 3 that, if Q_c is sufficiently high such that $Q_c \geq Q^*$, then the firm's optimal production quantity should be smaller than the capacity Q_c and the firm should sell its unused carbon credits under the cap-and-trade policy. For this case, the trade-off between reducing the production quantity and selling unused carbon credits is that the revenue reduction generated by decreasing Q from Q^* to Q_3^* should be compensated by selling the increments in the unused carbon credits (i.e., $Q^* - Q_3^*$).

Remark 2. Theorem 3 indicates that the firm's carbon emissions could be reduced when a *proper* cap-and-trade policy is implemented. Specifically, the amount of the carbon-emissions reduction depends on the values of α, β, c_u, and γ. In order to assure that the firm's carbon emissions are reduced to Q_c or less, the policy-maker should set the unit carbon-credit purchasing cost α no less than γ, i.e., $\alpha \geq \gamma$; otherwise, Policy 3 may not be effective in reducing carbon emissions that are generated by the firm.

We find from Theorem 3 that $Q_3^* = 0$ when $\beta \geq c_u$. This implies that the firm can profit more from selling carbon credits than from selling perishable products, when the price for carbon credits is *extremely* high. In practice, the policy-maker should effectively "manage" the cap-and-trade market to prevent the firm from acting as a carbon credit "dealer" instead of as a product "manufacturer."

Corollary 1. *When $Q^* > Q_c$ and $c_u > \beta > \gamma$, we find that*

$$\begin{cases} J_3(Q_3^*) \geq J(Q^*), \text{ if } \beta \geq \beta_0 \equiv c_u - (c_u + c_o) F(2Q_c - Q^*); \\ J_3(Q_3^*) < J(Q^*), \text{ if } \beta < \beta_0. \end{cases}$$

Proof. For a proof of this corollary, see 13.5.

From the above corollary, we note that, if $Q^* > Q_c$, $c_u > \beta > \gamma$, and $\beta \geq \beta_0$, then, as a result of implementing Policy 3, the firm's profit is increased (i.e., $J_3(Q_3^*) \geq J(Q^*)$) and its carbon emissions are decreased (i.e., $Q_3^* < Q_c$). That is, under the conditions that $Q^* > Q_c$, $c_u > \beta > \gamma$, and $\beta \geq \beta_0$, the firm should be willing to reduce its production quantity under Policy 3 and the policy is thus effective.

13.4 Numerical Study

In this section, we provide numerical examples to illustrate our analysis in Sect. 13.3. Since the analysis under Policy 1—which is provided in Sect. 13.3.1—is simple, we next compute the firm's optimal production quantities and expected profits under Policy 2 (carbon emissions tax) and Policy 3 (cap-and-trade). For simplicity, we assume that the firm does not incur a shortage cost (i.e., $s = 0$) and does not have a salvage value (i.e., $v = 0$). In addition, $X \sim N(500, 150)$, and $p = 100$. We consider several scenarios that differ in the values of other parameters including the unit acquisition cost c, the carbon tax τ, the unit carbon-credit purchasing cost α, the unit carbon-credit selling price β, and the prescribed emissions capacity Q_{c}.

13.4.1 Numerical Example for Policy 2

We now provide an example to illustrate our analysis for Policy 2 in Sect. 13.3.2. In this example, we use four different values of the unit cost c to represent four types of products, which include two high-profit products ($c = 15$ and $c = 35$) and two low-profit products ($c = 65$ and $c = 85$). For each product, we consider three scenarios, and for each scenario, we compute the corresponding optimal quantity for the firm.

In the first scenario, we assume that there is no capacity constraint. Accordingly, we calculate Q^* and $J(Q^*)$. In the second scenario, we assume that the carbon tax τ is equal to 10, and we calculate Q_2^* and $J_2(Q_2^*)$, which are then compared with Q^* and $J(Q^*)$ in the first scenario, respectively. We also compute the emissions reduction percentage $\kappa = (Q^* - Q_2^*)/Q^*$ and find the profit decrease percentage $\omega \equiv [J(Q^*) - J_2(Q_2^*)]/J(Q^*)$. In the third scenario, assuming that the firm desires to reduce carbon emissions by a specific percentage κ (e.g., $\kappa = 10\%$), we calculate Q_2^*, $J_2(Q_2^*)$, and ω; and also compute the corresponding tax rate τ in order to achieve the emissions reduction percentage κ. Our numerical results are presented in Table 13.4.

As Table 13.4 indicates, the firm's optimal production quantity is reduced as a result of implementing the carbon tax policy. From Scenario 2, we find that, if the per unit tax rate is 10, then the carbon emissions reduction for the high-profit products decreases as the profit margin $(p - c)$ decreases, whereas the reduction for the low-profit products significantly increases (from 9.73% to 26.38%) as the profit margin declines. We also note that the profit reduction percentage ω is strictly increasing in c; that is, if the profit margin is reduced, then the profit reduction percentage is increased.

In Scenario 3, when the carbon-emissions reduction percentage κ is equal to 10% for all products, the tax rate τ imposed on the high-profit product with $c = 35$ should be higher than that imposed on the high-profit product with $c = 15$. On the other hand, for the two low-profit products, the tax rate τ should be higher for the

Table 13.4 The firm's optimal quantities and corresponding expected profits in three scenarios

	High-profit		Low-profit	
c	15	35	65	85
Scenario 1: No carbon emissions policy				
Q^*	656	558	442	345
$J(Q^*)$	39,009	26,954	11,958	4,017
Scenario 2: Policy 2 with $\tau = 10$				
Q_2^*	601	519	399	254
$J_2(Q_2^*)$	32,741	21,377	7,748	965
$\kappa(\%)$	8.34	6.99	9.73	26.38
$\omega(\%)$	16.07	20.69	35.21	75.98
Scenario 3: Policy 3 with $\kappa = 10\%$				
Q_3^*	590	502	398	310
$J_2(Q_3^*)$	31,254	19,279	7,688	2,443
$\omega(\%)$	19.88	28.47	35.71	39.18
τ	12.5	14.5	10.2	4.8

Table 13.5 The numerical results when $Q^* \leq Q_c$

	High-profit product		Low-profit product	
c	15	35	65	85
c_u	85	65	35	85
Q_c	706	608	492	395
β	10	10	10	10
Q_3^*	601	519	399	254
$J_3(Q_3^*)$	39,799	27,652	12,666	4,914
$\kappa(\%)$	8.34	6.99	9.73	26.38
$\omega(\%)$	−2.03	−2.59	−5.92	−22.33

product with a smaller value of c. We also find that, even though the profit reduction percentage ω increases as the profit margin decreases, the increases for the four products are not as significant as those in Scenario 2.

13.4.2 Numerical Example for Policy 3

We now consider two examples to illustrate our analysis for Policy 3 in Sect. 13.3.3. From Theorem 3, we find that the firm's optimal quantity decision depends on the comparison between Q^* and Q_c. Next, we first present an example for the case that $Q_c \geq Q^*$, using the values of the unit acquisition cost c for four products as in Sect. 13.4.1. Setting the specific values of Q_c and β for each product, we present our calculation results in Table 13.5, where we find that, for each product, carbon emissions are decreased but the firm's expected profit is increased.

Next, we present another example to illustrate our analysis for the case that $Q_c < Q^*$. We set $\alpha = 12.5$ and $\beta = 10$, and we select three different values of Q_c for each product, as given in Table 13.6, where we find the following results.

Table 13.6 The numerical results when $Q^* > Q_c$

	High-profit product						Low-profit product					
c	15			35			65			85		
c_u	85			65			35			15		
Q^*	656			558			442			345		
$J(Q^*)$	39,009			26,954			11,958			4,017		
α	12.5			12.5			12.5			12.5		
β	10			10			10			10		
Q_c	640	595	550	550	515	480	435	390	345	340	230	180
β_0	5.42	26.01	49.53	4.01	22.44	39.36	3.57	21.09	30.42	1.45	14.53	14.98
γ	2.58	11.37	17.08	1.99	11.06	20.35	1.80	11.88	19.97	0.74	11.45	13.40
Q_3^*	601	595	590	519	515	510	399	390	386	254	230	207
$J_3(Q_3^*)$	39,139	38,685	38,125	27,072	26,720	26,287	12,096	11,637	11,076	4,363	3,246	2,632
$\kappa(\%)$	8.38	9.30	10.06	6.99	7.71	8.60	9.73	11.76	12.67	26.38	33.33	40.00
$\omega(\%)$	−0.33	0.83	2.27	−0.44	0.87	2.47	−1.15	2.68	7.38	−8.61	19.19	34.48

For each product, Q_3^* is reduced as Q_c is smaller; and, $J_3(Q_3^*)$ is greater than $J(Q^*)$ as long as $\beta > \beta_0$. We also find that Q_c more significantly impacts Q_3^* and $J_3(Q_3^*)$ for the low-profit products than for the high-profit products. In addition, if the profit margin is lower, then the impact of the carbon capacity on carbon emissions and the firm's expected profit are more significant.

13.5 Summary and Concluding Remarks

In this paper, we investigated the single-period problem under three carbon emissions policies including the mandatory carbon emissions capacity, the carbon emissions tax, and the cap-and-trade system. Under each policy, we obtained the optimal production quantity and calculated the corresponding expected profits for the firm. From our analysis, we draw some important *analytic* managerial insights. For example, we showed that, in order to reduce carbon emissions by a certain percentage, the tax rate τ imposed on the high-margin firm should be less than that on the low-margin firm for the high-profit perishable products, whereas the high-margin firm should absorb a higher tax than the low-margin firm for the low-profit products.

We also found that, from the perspective of the policy-maker, the emissions capacity should be set to a level such that the marginal profit of the firm is less than the carbon credit purchasing price, because, otherwise, the firm would produce more than the emissions capacity. We also derived the specific conditions under which, as a result of implementing the cap-and-trade policy, the firm's expected profit is increased and carbon emissions are reduced. The conditions assure the firm's and the policy-maker's incentives on the cap-and-trade policy.

The research problem discussed in this paper could be extended in several directions. In future, we may relax the single-period assumption and consider the quantity decisions of nonperishable products in multiple periods. In another possible research direction, we may also consider pricing decision for the firm, assuming the price-dependent aggregate demand in an additive and a multiplicative function form. In addition, from the policy-maker prospective, it would be nice if one could propose a way for a firm to select the best policy. The method of choosing the best carbon emission reduction policy for a given managerial situation likely has critical business implications for manufacturers.

Appendix A: Proofs of Theorems

Proof of Theorem 1. Temporarily ignoring the constraint that $Q \leq Q_c$, we can solve the classical single-period problem to find that

$$F(Q^*) = w \equiv \frac{p+s-c}{p+s-v}. \tag{13.8}$$

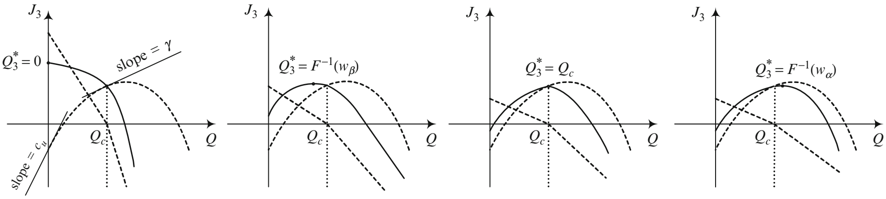


Fig. 13.1 The analysis of $J_3(Q)$ in four scenarios: (1) $\beta > c_u$, (2) $c_u > \beta > \gamma$, (3) $\beta < \gamma < \alpha$, and (4) $\alpha < \gamma$

Taking the constraint into consideration, we can easily obtain the result in this theorem. □

Proof of Theorem 2. To discuss the impact of w on the effectiveness of the carbon tax policy, we assume that the unit cost c in (13.4) takes two different values, e.g., c_1 and c_2 (w.l.o.g., $c_1 < c_2$); and then, ceteris paribus, the corresponding optimal quantities given by (13.2) are $\hat{Q}_2^*$ and $\tilde{Q}_2^*$, respectively.

Using (13.2) and (13.4), we find that, replacing c with $c + \tau$, the optimal production quantity is changed from Q^* to Q_2^*. If $\tau \to 0^+$, then $Q^* - Q_2^* = -\mathrm{d}Q^*/\mathrm{d}c$. Differentiating both sides of (13.8) once w.r.t. c, we have, $\mathrm{d}Q^*/\mathrm{d}c = -1/[(p+s-v)f(Q^*)]$. It thus follows that, as $\tau \to 0^+$, $\kappa = (Q^* - Q_2^*)/Q^* = 1/[(p+s-v)f(Q^*)]$, which is easily shown to be strictly increasing in Q^* when $Q^* \geq \mu$ but strictly decreasing in Q^* when $Q^* < \mu$. Therefore, for a high-profit product, $\hat{Q}_2^* > \tilde{Q}_2^*$, and $\hat{\kappa} \equiv (\hat{Q}^* - \hat{Q}_2^*)/\hat{Q}^* > \tilde{\kappa} \equiv (\tilde{Q}^* - \tilde{Q}_2^*)/\tilde{Q}^*$, whereas, for a low-profit product, $\hat{\kappa} < \tilde{\kappa}$. This theorem is thus proved. □

Proof of Theorem 3. We find from (13.6) that $J_3(Q)$ is a continuous, piecewise function in Q. We next consider two cases: $Q_c < Q^*$ and $Q_c \geq Q^*$; and for each case, we compute the corresponding optimal decision Q_3^*.

When $Q_c < Q^*$, we depict four scenarios as shown in Fig. 13.1; and, for each scenario, we compute the optimal solution Q_3^* as follows: If $\beta \geq c_u$, then we find from Fig. 13.1(1) that $J_3(Q)$ is strictly decreasing in Q over $[0, +\infty)$; and thus, the optimal quantity maximizing J_3 is $Q_3^* = 0$, and $J_3(Q_3^*) = \beta Q_c - s\mu$. If $c_u > \beta > \gamma$, then as Fig. 13.1(2) indicates, Q_3^* can be obtained as $Q_3^* = F^{-1}(w_\beta)$, which is in the range $(0, Q_c)$. If $\beta \leq \gamma \leq \alpha$, then, as Fig. 13.1(3) indicates, $J_3(Q)$ is increasing in $Q \in [0, Q_c]$ but decreasing in $Q \in (Q_c, +\infty)$. The optimal solution Q_3^* is thus determined as $Q_3^* = Q_c$. If $\alpha < \gamma$, then $Q_3^* = F^{-1}(w_\alpha) \in (Q_c, +\infty)$, as shown in Fig. 13.1(4).

When $Q_c \geq Q$, we find from (13.6) that $J_3(Q) = J(Q) + \beta(Q_c - Q)$, which is a concave function of Q. Similarly, we can show that $J_3(Q)$ is a decreasing, concave function of Q in the range $(Q_c, +\infty)$. Thus, the optimal solution Q_3^* must exist in the range $[0, Q_c]$. If $J_3(Q)$ is also strictly decreasing in $Q \in [0, Q_c]$, then $Q_3^* = 0$. Otherwise, Q_3^* should be obtained by solving $\mathrm{d}J_3(Q)/\mathrm{d}Q = 0$; that is, $Q_3^* = F^{-1}(w_\beta)$. Noting that $\mathrm{d}J_3(Q)/\mathrm{d}Q|_{Q=0} < 0$ *only* if $\beta > c_u$, we find that $Q_3^* = 0$

if $\beta > c_{\mathrm{u}}$; $Q_3^* = F^{-1}(w_\beta)$ otherwise. In addition, $Q_3^* \le Q_{\mathrm{c}}$ because $w_\beta \le w$; and, $J_3(Q_3^*) \ge J(Q^*) + \beta(Q_{\mathrm{c}} - Q^*) \ge J(Q^*)$. □

Appendix B: Proof of Corollary 1

We learn from Theorem 3 that, if $c_{\mathrm{u}} > \beta > \gamma$, then $Q_3^* = F^{-1}(w_\beta)$ and $\Phi(z_3^*) = w_\beta$. Hence, z_3^* is dependent on β, and $\phi(z_3^*)$ can be written as $\phi(z_3^*) = -[1/(p+s-v)] \times (\mathrm{d}\beta/\mathrm{d}z_3^*)$. Using (13.3), we have,

$$J_3(Q_3^*) - J(Q^*) = \beta(Q_{\mathrm{c}} - \mu) + \sigma(c_{\mathrm{u}} + c_{\mathrm{o}})[\phi(z^*) - \phi(z_3^*)]$$
$$= \beta(Q_{\mathrm{c}} - \mu) + \sigma\beta' + \sigma(c_{\mathrm{u}} + c_{\mathrm{o}})\phi(z^*). \quad (13.9)$$

Equating $J_3(Q_3^*)$ to $J(Q^*)$ and solving the resulting equation for β, we find that

$$\beta = \frac{\sigma(c_{\mathrm{u}} + c_{\mathrm{o}})\phi(z^*)}{Q_{\mathrm{c}} - \mu}[e^{(Q_{\mathrm{c}} - \mu)(z^* - z_3^*)/\sigma} - 1]. \quad (13.10)$$

Substituting β in (13.9) into (13.10), we obtain z_3^* as $z_3^* = z^* = 2(Q_{\mathrm{c}} - \mu)/\sigma - z^*$. It is easy to show that the corresponding value of β for z_3^* is $\beta_0 = c_{\mathrm{u}} - (c_{\mathrm{u}} + c_{\mathrm{o}})F(2Q_{\mathrm{c}} - Q^*)$. We also find that $J_3(Q_3^*) - J(Q^*) > 0$ for $\beta > \beta_0$, but $J_3(Q_3^*) - J(Q^*) < 0$ for $\beta < \beta_0$.

Acknowledgements The authors are grateful to two anonymous referees for their insightful comments that helped improve the paper. The Research of Mingming LENG was supported by the Research and Postgraduate Studies Committee of Lingnan University under Research Project No. DR09A3.

References

Accenture.com (2009). *Only one in 10 companies actively manage their supply chain carbon footprints*. http://newsroom.accenture.com/article_display.cfm?article_id=4801. Accessed November 18, 2011.

Benjaafar, S., Li, Y., & Daskin, M. (2010). *Carbon footprint and the management of supply chains: Insights from simple models*. http://www.ie.umn.edu/faculty/faculty/pdf/beyada-3-31-10.pdf. Accessed November 18, 2011.

BuySmart network (2008). *Reducing supply chain carbon backgrounder*. http://www.buysmartbc.com/cgi/page.cgi?aid=134&_id=76&zine=show. Accessed Accessed November 18, 2011.

Cachon, G. P. (2009). Carbon footprint and the management of supply chains. The INFORMS Annual Meeting, San Diego.

CO2List.org (2006). Co_2 released when making & using products. http://www.co2list.org/files/carbon.htm. Accessed November 18, 2011.

Congress of the United States (2008). *Policy options for reducing co_2 emissions: a CBO study*. This study was provided by the Congressional Budget Office (CBO).

Environmental News Energies Correspondent (2009). *How to have a low-carbon Christmas*. http://www.enviro-news.com/news/how_to_have_a_lowcarbon_christmas.html Accessed November 18, 2011.

Hadley, G., & Whitin, T. M. (1963). *Analysis of inventory systems*. Englewood Cliffs: Prentice-Hall.

Hoen, K. M. R., Tan, T., & Fransoo, J. C. (2010). *Effect of carbon emission regulations on transport mode selection in supply chains*. http://cms.ieis.tue.nl/Beta/Files/WorkingPapers/Beta_wp308.pdf. Accessed November 18, 2011.

Hua, G., Cheng, T. C. E., & Wang, S. (2010). *Managing carbon footprints in inventory control*. Working paper.

International Energy Agency (2008). *Carbon capture and storage (CCS) for power generation and industry*. D:\Temp\NewReferences\IEA-Technology Roadmaps CCS for Power Generation and Industry.htm. Accessed November 18, 2011.

Khouja, M. (1999). The single-period (newsvendor) problem: literature review and suggestions for future research. *Omega*, *27*, 537–553.

Kim, N. S., Janic, M., & Wee, B. (2009). Trade-off between carbon dioxide emissions and logistics costs based on multiobjective optimization. *Transportation Research Record*, *2139*, 107–116.

Letmathe, P., & Balakrishnan, N. (2005). Environmental considerations on the optimal product mix. *European Journal of Operational Research*, *167*, 398–412.

Penkuhn, T., Spengler, T., Püchert, H., & Rentz, O. (1997). Environmental integrated production planning for ammonia synthesis. *European Journal of Operational Research*, *27*, 327–336.

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

The Huber Group (2003). *Newspaper inks and the environment*. Technical information of the Huber Group.

Chapter 14
Optimal Decisions of the Manufacturer and Distributor in a Fresh Product Supply Chain Involving Long-Distance Transportation

Xiaoqiang Cai, Jian Chen, Yongbo Xiao, Xiaolin Xu, and Gang Yu

Abstract We consider a supply chain in which a distributor procures from a manufacturer a type of fresh product, which has to undergo long distance transportation before reaching the market. In addition to the risk caused by random fluctuations of the market demand, the distributor also faces the risk that the product procured may decay and deteriorate during transportation. The market demand for the product depends on its level of freshness and the distributor's selling price. The manufacturer has to determine his wholesale price based on its effect on the order quantity of the distributor, whereas the distributor has to determine his order quantity and selling price, based on the wholesale price, the likely loss of the product in long distance transportation, the product's level of freshness when it reaches the market, and the possible demand for the product. We develop a model to formulate this problem, and derive each party's optimal decisions in both uncoordinated and coordinated situations. We introduce a new incentive scheme to facilitate the coordination of the two parties, which comprises two parts: (1) the

X. Cai (✉)
Department of Systems Engineering & Engineering Management,
The Chinese University of Hong Kong, Shatin, N.T., Hong Kong
e-mail: xqcai@se.cuhk.edu.hk

J. Chen • Y. Xiao
Research Center for Contemporary Management, Key Research Institute
of Humanities and Social Sciences at Universities, School of Economics and Management,
Tsinghua University, Beijing 100084, China
e-mail: chenj@sem.tsinghua.edu.cn; xiaoyb@sem.tsinghua.edu.cn

X. Xu
School of Business, Nanjing University, Nanjing 210093, China
e-mail: xuxl@nju.edu.cn

G. Yu
Department of Management Science and Information Systems,
The University of Texas at Austin, Austin, TX, USA
e-mail: yu@uts.cc.utexas.edu

T.-M. Choi (ed.), *Handbook of Newsvendor Problems*, International Series in Operations Research & Management Science 176, DOI 10.1007/978-1-4614-3600-3_14,

manufacturer offers his wholesale price as a function of the actual transportation time and price discount in the market; and (2) the manufacturer compensates the distributor for any unsold unit of the product. We show that this incentive scheme can induce the distributor to order up to the quantity required to maximize the total benefit of the centralized system, and both parties will all be better off than in the uncoordinated case. Computational studies are also conducted, which reveal some interesting managerial insights.

Keywords Fresh product supply chain • Long distance transportation • Coordination • Level of freshness • Incentive scheme

14.1 Introduction

The production of food products, such as live seafood, fresh fruit, fresh vegetables, etc., is highly characterized by geographical location. As such, it is quite common that a product which is produced in one region is sold to another distant market in which the product is consumed. For example, fresh fruit is imported into China from California, fresh vegetables are exported from China to other countries such as Europe and Japan, live seafood is imported into Hong Kong from Australia, Canada, and the United States of America. According to OECD (2005), agricultural goods trade has become a significant portion of the GDP of many countries: in 2003, EU-15 and U.S. exports of agriculture accounted for 8.4% and 8.6% of their GDPs, respectively, whereas that for New Zealand reached 48.1%. The monthly statistics of the Commerce Ministry of China (December, 2005) show that in 2005, China's exports of fishery and seafood products totaled more than $4.35 billion, and exports of vegetables reached $3.05 billion. According to U.S. Horticulture Exports to Asia (Sanchez and Kipe 2004), "In the calendar year (CY) 2003, U.S. horticultural exports to the world totaled more than $12.3 billion. Approximately 28 percent, or $3.5 billion, of these exports were shipped to markets of Asia. Asia is a significant market for U.S. horticultural products, with 10 of the top 25 horticultural export markets located in the region." In addition to international trading, in countries of vast geographical size, it is also common that fresh products are produced in one place and sold in another domestic, but distant market.

Because of the geographical difference between the origin of production and the destination of consumption, a common feature of the supply chain of such products is the requirement of long distance transportation. For example, according to Hagen et al. (1999), "In California more than 485,000 truckloads of fresh fruits and vegetables travel 100 to 3,100 miles to reach their destinations." While there is a great profit potential in selling a product in a market in which the product is not readily producible, long distance transportation of fresh products involves, however, a high risk. Due to the perishable nature of the products and the transport time that is required, it is almost inevitable that a certain degree of decay or deterioration will have taken place by the time the products arrive at the distant markets. As indicated by Kasmire (1999): "All fresh products continue to deteriorate with time, even under optimum handling and transport conditions. Postharvest and transport times should be kept as short as possible, especially under less than optimum conditions, to limit deterioration and extend marketable life of products." Long distance transportation is also prone to unexpected delays, which may be caused by disruptive variations of weather, equipment breakdowns, transport network congestion, etc. When such a delay occurs, the products may experience even severer decay and deterioration during the prolonged transportation process.

The market demand for a fresh product, however, often depends on the product's level of freshness and the selling price. How to determine the appropriate pricing decision, by considering the freshness of the product and its correlation with the market demand? This is an important question facing all fresh product distrubtors/retailers. Before (and in addition to) this pricing decision, the distributor has another critical decision to make; i.e., the quantity of the product to be procured from the manufacturer, based on the wholesale price charged by the manufacturer and all other information including the transportation time, the possible loss in the order quantity due to decay, and the market demand. The manufacturer, on the other hand, has to determine the wholesale price based on its possible effect to the order quantity of the distributor. How can we model the decisions to be made by the manufacturer and the distributor? What are their optimal decisions if they do not coordinate and want to optimize their individual objectives? What are their decisions to optimize the joint objective of the centralized system? Can any incentive schemes (contracts) be introduced to facilitate the coordination between the two parties, so that the risks involved in the supply chain can be shared with a mutually acceptable mechanism to allow both parties to be all benefited by the coordination? In particular, because time is a crucial element for perishable products, how should decisions be made with consideration of the relationship between perishability and time?

The main purpose of our research is to investigate the above questions. Specifically, we will consider a supply chain of fresh products, with the following features: (1) Uncertain transportation time; (2) decay and deterioration during transportation, which may cause reduction in both quality (freshness level) and quantity, respectively; and (3) random market demand, sensitive to the selling price and the level of freshness of the product when it reaches the market. We obtain the following results:

(1) The optimal decisions for both the manufacturer and the distributor in the decentralized system, to optimize their respective objectives;
(2) The optimal decisions in the centralized system, to optimize their joint objective;
(3) A transportation-time dependent, price-discount sharing mechanism and a compensation scheme, which can ensure that both parties take the coordinated decisions and are all better off by the coordination; and
(4) Managerial insights, uncovered from both theoretical results and computational studies.

The remainder of the paper is organized as follows. In Sect. 14.2, we will provide a brief review of the related literature. Problem formulation, assumptions, and notation will be presented in Sect. 14.3. In Sect. 14.4, we will derive the optimal decisions of the manufacturer and the distributor in the uncoordinated, decentralized system. Optimal decisions for the centralized system will be developed in Sect. 14.5, followed by the development of an incentive scheme that motivates both parties to adopt the coordinated decisions in Sect. 14.6. Sect. 14.7 will report and analyze the numerical results obtained from our computational studies. Finally, concluding remarks will be given in Sect. 14.8.

14.2 Related Literature

Studies on supply chain management of perishable products have begun with concerns about inventory management of such products, which is to analyze and determine the replenishment policies for inventory. An early work on a perishable inventory problem was described by Whitin (1957), where fashion goods deteriorating at the end of certain storage periods were considered. Since then, considerable attention has been paid to this line of research. Nahmias (1982) provides a comprehensive survey of research published before the 1980s. More recent studies on the deteriorating inventory models can be found in Raafat (1991), Goyal and Giri (2001), and Karaesmen et al. (2011), in which relevant literature published in 1980s, 1990s, and 2000s are reviewed respectively.

Our model considers fresh products that may face the two kinds of loss: decreases in quantity and quality. That is, as time goes by, some quantity of the product procured becomes obsolete (for example, some of the fish die and become valueless); Meanwhile, the product also becomes less fresh gradually. Our model therefore deals with both the quantity and quality effects: the quantity decrease affects the effective product supply, whereas the quality (or freshness level) deterioration affects the market demand. Both effects will be captured in our model, by using general functions based on the elapsed transportation time. The work of Rajan et al. (1992) is an earlier study considering both value drop and quantity decrease. Their focus is, however, on the inventory replenishment and pricing decisions of perishable products. They study a model in which the demand is deterministic, and the decision maker optimizes the price $p(t)$ and the order cycle length to maximize the average profit per unit time (the optimal price is assumed to be a deterministic function of t).

We investigate the coordination between the manufacturer and the distributor, which takes into account the decay and deterioration of the product during transportation and in the market. Coordination between suppliers and distributors (or retailers) has been a subject of extensive study in supply chain management in the last few decades (see, e.g., Chen 2003; Song et al. 2005). Such coordination is usually achieved through "contracts" between the upstream suppliers and the downstream distributors, to increase the total supply chain profit so as to make it closer to the profit that can be generated from a centralized control (channel coordination), or to share the risks among the supply chain partners (Tsay et al.

1999). Various models of supply chain contracts have been developed in the literature. Price discount is often suggested as an incentive to facilitate coordination (see Parlar and Wang 1994; Weng 1995; Wang 2001). Other incentive schemes include quantity commitment (Anupindi and Bassok 1999), quantity flexibility contracts (Tsay 1999), backup agreements (Eppen and Iyer 1997), buy back or return policies (Pasternack 1985), revenue sharing (Cachon and Lariviere 2005), sales rebate or "markdown allowance" (Taylor 2000; Krishan et al. 2004). Cachon (2003) provides an excellent survey on supply chain coordination with contracts. As Cachon (2003) points out, the buy backs, quantity flexibility, and sales-rebate contracts do not coordinate in settings in which the demands are price dependent. These contracts encounter problems because the incentives they provide to coordinate the distributor's quantity action distorts the pricing decision. One of the main contributions of our work is the design of a price-discount sharing mechanism (which is also time dependent) together with a compensation scheme to coordinate the manufacturer and the distributor.

By considering long distance transportation with uncertain transport time, we actually take the uncertain lead-time into account. In the production-inventory literature, there are numerous papers published that investigate the stochastic lead times (see, e.g., Eppen and Martin 1988; Lau and Zhao 1993; and Bookbinder and Cakanyildirim 1999). Most of them study multi-period models and focus on the influence of the stochastic lead time upon the inventory/safety inventory. Under a discrete multi-period inventory setting, Lodree and Uzochukwu (2008) and Shen et al. (2010) consider time-dependent demands of perishable product. Our model, however, deals with the effects of stochastic lead time on the effective inventory and demand simultaneously. In a broader sense, our work represents a new attempt on supply chain coordination considering the effects of uncertain lead times.

Our model addresses the issue that the distributor is risk averse towards the loss that may be incurred by the random transportation time. By use a parameter $\rho \geq 0$, our model considers the different degree of risk aversion. While the majority of the literature on supply chain coordination is based on the assumption that decision makers are risk neutral (consequently, the objective commonly considered is to maximize the expected profit), there have been some studies that examine the influence of risk aversion in supply chain contracting (see, e.g., Agrawal and Seshadri 2000; Plambeck and Zenios 2000). The conventional way of modeling risk attitude is to set the objective as maximizing the expected utility instead of the expected profit. An exponential utility function has been used by Feng and Xiao (2008). Another approach is to set a target revenue (or profit) level, and to apply a "downside penalty" if the target level is not reached (Nawrocki 1999). A "required probability of achieving at least the maximum expected profit" can be imposed as a constraint of the optimization problem (Weng 1999). Our model also uses "downside penalty" to reflect the risk-averse attitude of the distributor on the uncertain transportation delays. That is, in cases when the realized profit does not reach its expected level, a penalty will be incurred, the magnitude of which is proportional to the profit gap relative to the expected profit.

Our model can be regarded as an extension of the traditional newsvendor problem (of one supplier and one distributor) with price-dependent demands. Our model differs, however, from the previous literature in two aspects. First, the order quantity and selling price decisions, which can be made simultaneously in the newsvendor models, are determined sequentially (and optimally) because the ordering decision has to be made before the product is ready for transportation, whereas the optimal selling price, which depends on the freshness level of the product and the timing of the market, can be determined only when the product arrives at the distributor. That is, the optimal selling price depends on the arrival time, which is not known at the point when the order quantity is determined. Second, the contract to facilitate the coordination between the manufacturer and the distributor may also correlate to the (stochastic) transportation time.

Recently, Ray et al. (2005) have studied a problem to determine the optimal pricing and stocking decisions, in which the market demand is stochastic and price sensitive, and the delivery times are random. In addition to addressing these important issues in our model, we investigate, moreover, the specific requirements imposed by fresh and perishable products. The optimal uncoordinated and coordinated decisions and strategies we derive for the fresh-product supply chain depend significantly on the features of decay and deterioration of the product. The pricing decision, for example, in our model has to be made based on the level of freshness of the product when it reaches the market and the effective supply that remains after long distance transportation. Under a risk-neutral setting, Cai et al. (2010) study the optimization and coordination of a fresh product supply chain, where the freshness-keeping effort of the downstream distributor is taken into account as a decision variable. Our present paper differs from Cai et al. (2010) mainly in two aspects: (1) We consider a risk-averse setting, under which the degree of risk aversion of the distributor affects his decisions; and (2) rather than assuming that the transportation time is an endogenous controllable variable, we consider the situation in which it is an exogenous random variable.

14.3 The Model

We investigate the following problem. A distributor procures a kind of fresh product from a manufacturer on a make-to-order basis. The product ordered has to undergo long distance transportation before it reaches the market. The transaction between the manufacturer and the distributor is conducted on a free-on-board (FOB) basis, under which the price of the seller (manufacturer) includes the cost of loading onto the transport vessel at the designated point, whereas the transportation and insurance are the responsibilities of the buyer (distributor).

The product is fully fresh when it is loaded onto the vehicle (e.g., the cargo ship). It remains fresh during a period that we call its fresh duration τ. The fresh duration τ, where $\tau \geq 0$, is a constant which depends on the nature of the product and the way to treat and keep it (Kasmire 1999). After that, the product starts to perish at a significant rate. The perishability may lead to "deterioration" or "obsolescence," which may all occur during the process of transportation. Deterioration lessens the quality (freshness) of the product, whereas obsolescence reduces its effective (marketable) quantity. Specifically, we model the two types of perishability by the following two time-dependent indices, where $t = 0$ is the time when the product is uploaded to the vehicle:

- A function $\theta(t)$ of time t, defined over $[0,1]$, as the freshness index of the product. $\theta(t) = 1$ if $t \leq \tau$ and $0 \leq \theta(t) < 1$ otherwise. Especially, if the freshness level declines at a constant exponential rate δ, then $\theta(t)$ takes the following form:

$$\theta(t) = \begin{cases} 1 & \text{if } t \leq \tau; \\ \exp(-\delta(t-\tau)) & \text{otherwise.} \end{cases}$$

- A function $m(t)$ of time t, define over $[0,1]$, as the index on the marketable quantity of the product at time t. Suppose that q units of the product are loaded to the vehicle, the amount becomes $qm(t)$ after a period of time t, where $0 < m(t) \leq 1$.

Note that exponential functions have been used in the literature to model the quantity decrease and quality decline of perishable products; see Raafat (1991) and Rajan et al. (1992). Our functions $\theta(t)$ and $m(t)$ are not limited to the exponential forms, which can be any functions depending on the nature of the product.

The market demand for the product depends on its freshness level $\theta(t)$ when it reaches the market at time t and the selling price p of the distributor, with the following multiplicative functional-form:

$$D(p,t) = y_0 \varepsilon p^{-k} \theta(t), \quad k > 1,$$

where y_0 is a constant, ε is a random variable, and k is the price-elasticity index. The variable ε reflects the random fluctuations of the market demand. We assume that its PDF and CDF are $f(x)$ and $F(x)$, respectively. Without loss of generality, we assume that the mean of ε is equal to 1 (this can be achieved by appropriately scaling the parameter y_0). The parameter k reflects the sensitivity of the demand to the price. The larger the k value, the more sensitive the demand to the price. If $k > 1$, then the demand is price-elastic, and inelastic otherwise. We focus on the price elastic case (If $k \leq 1$, then we can show that the optimal price under consideration goes to infinity; see Remark 1.5 below). Note that if $\theta(t) = 1$ for all t (that is, the product is not deteriorative), then the demand function above reduces to that considered by Petruzzi and Dada (1999), Wang (2004), and Wang et al. (2004).

Denote T as the transportation time, which is a continuous random variable distributed over $[a,b]$, with its CDF and PDF being $G(t)$ and $g(t)$, respectively. When $b = a$, the model reduces to the special case with a deterministic (fixed) transportation time. Assume that the transportation cost is $c_{\mathrm{T}} q$ if the quantity of the product to be transported is q, where c_{T} is the unit transportation cost.

14.3.1 The Decision of the Manufacturer

Let c_{M} be the unit production cost of the manufacturer. In the situation where there is no coordination between the manufacturer and the distributor, the manufacturer is to determine a wholesale price w so as to maximize his profit π_{m}:

$$\pi_{\mathrm{m}}(w) = (w - c_{\mathrm{M}})q, \tag{14.1}$$

where q is the order quantity of the distributor, which is influenced by the whole sales price w.

14.3.2 The Decision of the Distributor

Let q and p be the order quantity and the selling price of the distributor. For any transportation time t, the expected profit of the distributor w.r.t. the random demand is given below:

$$\pi_{\mathrm{d}}(p,q|t) = pE_{\varepsilon}\{\min\{D(p,t), qm(t)\}\} - wq - c_{\mathrm{T}}q. \tag{14.2}$$

The profit function above has an implicit assumption that the salvage value of any unsold product is zero. This is common for fresh products as they usually become valueless if they cannot be sold within a period of time.

The distributor is risk averse to the transportation delay (which could incur a very substantial loss). Thus, we introduce the following utility function:

$$U_{\mathrm{d}}(p,q) = E_t\left\{\pi_{\mathrm{d}}(p,q|t) - \rho\left[E_t\{\pi_{\mathrm{d}}(p,q|t)\} - \pi_{\mathrm{d}}(p,q|t)\right]^+\right\}, \tag{14.3}$$

where $\rho \geq 0$ and x^+ is defined as $x^+ = \max\{0, x\}$. It is easy to see that $[E_t\{\pi_{\mathrm{d}}(p,q|t)\} - \pi_{\mathrm{d}}(p,q|t)]^+$ represents the downside penalty, which will occur if the realized profit is lower than the expected profit. Thus, ρ represents the degree of risk aversion. If $\rho = 0$, the utility reduces to the expected profit, that is, the risk-neutral case.

In the situation where there is no coordination between the manufacturer and the distributor, the distributor is to determine q and p so as to maximize $U_{\mathrm{d}}(p,q)$.

14.3.3 The Decisions of the Manufacturer and the Distributor Under a Joint Objective

In the centralized system, the manufacturer and the distributor should make decisions to optimize their joint objective. For any transportation time t, their joint expected profit with respect to the random demand is given below:

$$\begin{aligned}\Pi_{\mathrm{c}}(p,q|t) &= \pi_{\mathrm{m}}(w|t)+\pi_{\mathrm{d}}(p,q|t)\\ &= pE_{\varepsilon}\{\min\{D(p,t),qm(t)\}\}-(c_{\mathrm{M}}+c_{\mathrm{T}})q \end{aligned} \tag{14.4}$$

Note that the wholesale price w vanishes in the centralized system (in which the selling profit of the manufacturer equals the purchase cost of the distributor). Thus, the expected profit for the centralized system is equal to the revenue generated from the market minus the production and the transportation costs, as shown by (14.4) above.

The two parties should maximize their joint utility function:

$$U(p,q)=E_t\left\{\Pi_{\mathrm{c}}(p,q|t)-\rho\left[E_t\{\Pi_{\mathrm{c}}(p,q|t)\}-\Pi_{\mathrm{c}}(p,q|t)\right]^{+}\right\}, \tag{14.5}$$

where $\rho \geq 0$.

Again, the utility function reduces to the expected profit when $\rho = 0$.

14.4 Optimal Decisions in the Decentralized System

In this section, we will characterize the optimal decisions of the manufacturer and the distributor in the decentralized system in which they are not coordinated. We will do so by using a backward approach. First, we will derive the optimal selling price p^* of the distributor, given any *arbitrary* order quantity q, wholesale price w, and transportation time t. We will then derive his optimal order quantity q^* based on its relationship with p^*. The optimal wholesale price w^* of the manufacturer will be obtained based on its relationship with $q*$.

Following Petruzzi and Dada (1999), we define

$$z := qm(t)/[y_0 p^{-k}\theta(t)],$$

and call it the "stocking factor." Then, the problem of choosing a price p is equivalent to choosing a stocking factor z. By substituting $p=(zy_0\theta(t)/qm(t))^{1/k}$ into (14.2), the distributor's objective function can be rewritten as

$$\pi_{\mathrm{d}}(z|q,t)=\left(\frac{zy_0\theta(t)}{qm(t)}\right)^{1/k}\cdot E_{\varepsilon}\left\{\min\left\{\frac{qm(t)}{z}\varepsilon,qm(t)\right\}\right\}-(w+c_{\mathrm{T}})q. \tag{14.6}$$

Lemma 1. *The optimal stocking factor z that maximizes $\pi(z|q,t)$ is determined by*

$$\int_0^z (k-1)xf(x)\mathrm{d}x = z[1-F(z)]. \tag{14.7}$$

Moreover, if ε has a generalized increasing failure rate (GIFR), and $\lim_{x\to\infty} x[1-F(x)] = 0$*, then (14.7) has a unique solution* z_0*.*

Proof. Equation (14.6) can be transformed as follows:

$$\begin{aligned}\pi_{\mathrm{d}}(z|q,t) &= (zy_0\theta(t))^{1/k}(qm(t))^{1-1/k}\cdot E_\varepsilon\left\{\min\left\{\frac{\varepsilon}{z},1\right\}\right\} - (w+c_{\mathrm{T}})q \\ &= (zy_0\theta(t))^{1/k}(qm(t))^{1-1/k}\left(1-\int_0^z (1-x/z)f(x)\mathrm{d}x\right) - (w+c_{\mathrm{T}})q.\end{aligned} \tag{14.8}$$

The optimal stocking factor z that maximizes $\pi_{\mathrm{d}}(z|q,t)$ must satisfy the first order condition:

$$\frac{\mathrm{d}\pi_{\mathrm{d}}(z|q,t)}{\mathrm{d}z} = \frac{(y_0\theta(t))^{1/k}(qm(t))^{1-1/k}}{z^{1-1/k}k}\left(1-\int_0^z [\frac{x}{z}(k-1)+1]f(x)\mathrm{d}x\right). \tag{14.9}$$

Since $(y_0\theta(t))^{1/k}(qm(t))^{1-1/k}z^{1/k-1}k^{-1} > 0$, $\mathrm{d}\pi_{\mathrm{d}}(z|q,t)/\mathrm{d}z = 0$ implies

$$\int_0^z [\frac{x}{z}(k-1)+1]f(x)\mathrm{d}x = 1,$$

which gives (14.7).

We next prove the uniqueness of z_0. Let

$$\phi(z) := \int_0^z [x(k-1)+z]f(x)\mathrm{d}x - z = -z\bar{F}(z) + (k-1)\int_0^z xf(x)\mathrm{d}x,$$

where $\bar{F}(z) = 1-F(z)$. Then we have

$$\phi'(z) = \int_0^z f(x)\mathrm{d}x + zkf(z) - 1 = zkf(z) - \bar{F}(z) = k\bar{F}(z)\left[h(z)-\frac{1}{k}\right],$$

where $h(z)$ is the generalized failure rate function of ε, i.e., $h(z) = zf(z)/\bar{F}(z)$. When $h(z)$ increases in z, we know that $\phi(z)$ decreases before z reaches $h^{-1}(1/k)$ and increases after $h^{-1}(1/k)$, and hence is unimodal (see Fig. 14.1). Since $\phi(0) = 0$ and $\lim_{z\to\infty}\phi(z) > 0$, it is apparent that $\phi(z) = 0$ has only one solution within $(0,\infty)$. Therefore, z_0 is uniquely determined by (14.7). It is also clear that for $z > z_0$, $\phi(z) > 0$ and thus $\mathrm{d}\pi_{\mathrm{d}}(z|q,t)/\mathrm{d}z < 0$; for $0 < z < z_0$, $\phi(z) < 0$ and thus $\mathrm{d}\pi_{\mathrm{d}}(z|q,t)/\mathrm{d}z > 0$. Therefore, $\pi_{\mathrm{d}}(z|q,t)$ is also unimodal in z, and z_0 is the unique maximizer (instead of minimizer) of $\pi_{\mathrm{d}}(z|q,t)$. This completes the proof. □

Note that both GIFR and $\lim_{x\to\infty} x\bar{F}(x) = 0$ are very mild restrictions on the demand distribution. GIFR is a weaker condition than IFR (increasing failure rate)—a property that is known to be satisfied by distributions such as normal,

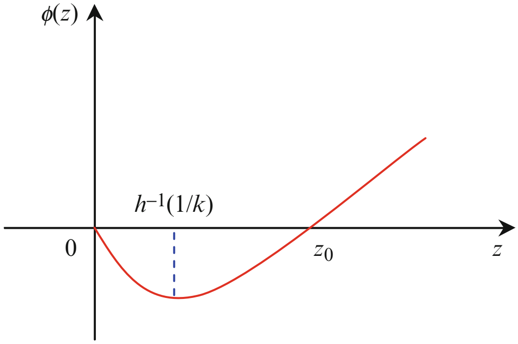


Fig. 14.1 The curve of function $\phi(z)$

uniform, and also the gamma and Weibull families, subject to parameter restrictions (Barlow and Proschan 1965). The condition $\lim_{x\to\infty} x\bar{F}(x) = 0$ is also satisfied by the above-mentioned distribution functions.

From (14.7), we can see that the optimal stocking factor is dependent on the price-elasticity k and the distribution of the random factor ε, but independent of other parameters (including the order quantity and the transportation time). The optimal selling price p^* can therefore be obtained directly. This is summarized in the theorem below.

Theorem 1. *Suppose that the distributor orders a quantity q and the product arrives at time t. Then the optimal selling price $p^*(q,t)$ of the distributor is:*

$$p^*(q,t) = \left(\frac{y_0 z_0 \theta(t)}{qm(t)}\right)^{1/k}. \tag{14.10}$$

Remark 1. (1) Note that $\theta(t)$ is the product's level of freshness when it reaches the market at time t. It follows from (14.10) that, the higher the level of freshness, the higher the selling price. Note also that $qm(t)$ is the marketable quantity of the product (effective supply) when it reaches the market at time t. It is easy to see from (14.10) that, the smaller the effective supply, the higher the price.

(2) It is interesting to analyze the relationship between the optimal selling price $p^*(q,t)$ and the transportation time t. A general intuition may suggest that the price may decrease if a transportation delay occurs (t is longer) because the product may have become less fresh. The key here is, nevertheless, that the selling price also depends on the effective supply—If a transportation delay occurs, the marketable quantity $qm(t)$ may have dropped and thus the effective supply to the market is reduced.

(3) Our result shows that whether $p^*(q,t)$ decreases or increases in t depends on the function $\theta(t)/m(t)$, which we call the "quality–quantity" ratio. If the quantity decreases quickly (a small $m(t)$) but the quality drops slowly (a large $\theta(t)$), then the optimal selling price may be even higher than the "normal" optimal selling price $(y_0 z_0/q)^{1/k}$ with a fully fresh product.

(4) Suppose that the actual transportation time is t and the order quantity of the distributor is q. Equation (14.10) gives a close-loop formula to determine the optimal price: When the product arrives at t, the distributor can observe "the level of freshness" and "the effective supply" and then set his selling price.
(5) If the price elasticity $k \leq 1$, from (9) we can see that $\frac{d\pi_{\rm d}(z|q,t)}{dz} > 0$ holds; that is, $\pi_{\rm d}(z|q,t)$ is strictly increasing in z. Therefore, the optimal stocking factor z_0 goes to infinity, and so is the distributor's optimal retail price.

We can now determine the optimal order quantity of the distributor, under any wholesale price of the manufacturer. By substituting $p^*(q,t)$ of (14.10) into (14.2), we can rewrite the distributor's optimal expected profit conditional on the transportation time t as

$$\begin{aligned}\pi_{\rm d}(q|t) :&= \pi_{\rm d}(p^*(q,t)|q,t)\\ &= \frac{k}{k-1}(y_0\theta(t))^{1/k}(qm(t))^{1-1/k}z_0^{1/k}[1-F(z_0)]-(w+c_{\rm T})q. \quad (14.11)\end{aligned}$$

As stated in Sect. 14.3, the distributor seeks to maximize his expected utility $U_{\rm d}(q)$:

$$U_{\rm d}(q) = E_t\{U_{\rm d}(q,t)\} = \pi_{\rm d}(q) - \rho E_t\{[\pi_{\rm d}(q) - \pi_{\rm d}(q|t)]^+\}, \quad (14.12)$$

where $\pi_{\rm d}(q) = E_t\{\pi_{\rm d}(q|t)\}$.

Theorem 2. *For any given wholesale price w of the manufacturer, the distributor's optimal order quantity should be:*

$$q^* = y_0 z_0\left(\frac{1-F(z_0)}{w+c_{\rm T}}[K_0 - \rho E_t\{S(t)^+\}]\right)^k, \quad (14.13)$$

where $K_0 = E_t\{\theta(t)^{1/k}m(t)^{1-1/k}\}$ and $S(t) = K_0 - \theta(t)^{1/k}m(t)^{1-1/k}$. The corresponding optimal expected utility value is:

$$U_{\rm d}^* = \frac{(w+c_{\rm T})y_0z_0}{k-1}\left(\frac{1-F(z_0)}{w+c_{\rm T}}[K_0 - \rho E_t\{S(t)^+\}]\right)^k. \quad (14.14)$$

Proof. Taking expectations with respect to t, we have:

$$\pi_{\rm d}(q) = E_t\{\pi_{\rm d}(q,t)\} = \frac{k}{k-1}y_0^{1/k}q^{1-1/k}z_0^{1/k}[1-F(z_0)]K_0 - (w+c_T)q. \quad (14.15)$$

Hence,

$$\pi_{\rm d}(q) - \pi_{\rm d}(q|t) = \frac{k}{k-1}y_0^{1/k}q^{1-1/k}z_0^{1/k}[1-F(z_0)]S(t), \quad (14.16)$$

where $S(t) = K_0 - \theta(t)^{1/k}m(t)^{1-1/k}$.

Therefore,

$$U_{\mathrm{d}}(q) = \frac{k}{k-1} y_0^{1/k} q^{1-1/k} z_0^{1/k} [1 - F(z_0)] \left(K_0 - \rho E_t\{S(t)^+\}\right) - (w + c_{\mathrm{T}})q. \quad (14.17)$$

Notice that when the downside penalty parameter ρ is sufficiently large, i.e., $\rho \geq K_0/E_t\{S(t)^+\}$, the distributor's utility function will be strictly decreasing in q and hence the optimal decision is $q^* = 0$. To avoid such trivial cases, in the following we limit ρ within the interval $[0, K_0/E_t\{S(t)^+\})$. The first and second order conditions are as follows:

$$\frac{\mathrm{d}U_{\mathrm{d}}(q)}{\mathrm{d}q} = -(w + c_{\mathrm{T}}) + y_0^{1/k} z_0^{1/k} [1 - F(z_0)] \left(K_0 - \rho E_t\{S(t)^+\}\right) q^{-1/k}, \quad (14.18)$$

$$\frac{\mathrm{d}^2 U_{\mathrm{d}}(q)}{\mathrm{d}q^2} = -\frac{1}{k} y_0^{1/k} z_0^{1/k} [1 - F(z_0)] \left(K_0 - \rho E_t\{S(t)^+\}\right) q^{-1-1/k} < 0. \quad (14.19)$$

Therefore, $U_{\mathrm{d}}(q)$ is concave in q, and the optimal order quantity q^* that maximizes $U_{\mathrm{d}}(q)$ is determined by the first-order condition (14.18), from which (14.13) is obtained. Substituting q^* of (14.13) into (14.17), we obtain the distributor's optimal utility, which is (14.14). □

For brevity, let us denote

$$K_1 := K_0 - \rho E_t\{S(t)^+\}. \quad (14.20)$$

From Theorem 2, we have the following observations.

Remark 2. (1) q^* decreases in the manufacturer's wholesale price w and the unit transportation cost c_{T}.

(2) We can also show that:

$$q^* = y_0 z_0 \left(\frac{k-1}{k} \cdot \frac{1 - F(z_0)}{c_{\mathrm{T}} + c_{\mathrm{M}}} K_1 \right)^k, \quad (14.21)$$

where

$$K_1 = K_0 - \rho E_t\{S(t)^+\} \leq K_0 < 1.$$

Should the transportation lead time be zero (or the product not be perishable), the distributor would have ordered

$$q^0 = y_0 z_0 \left(\frac{k-1}{k} \cdot \frac{1 - F(z_0)}{c_{\mathrm{T}} + c_{\mathrm{M}}} \right)^k. \quad (14.22)$$

Therefore, $q^* < q^0$. That is, the possibility of deterioration and decay discourages the distributor from ordering more.

We can now develop the optimal wholesale price w^* of the manufacturer, based on the relationship with q^* as defined by (14.13). The manufacturer's profit function, denoted by $\pi_{\text{m}}(w)$, can now be expressed as

$$\pi_{\text{m}}(w) = (w - c_{\text{M}})q^* = y_0 z_0 \left(\frac{1 - F(z_0)}{w + c_{\text{T}}} K_1 \right)^k (w - c_{\text{M}}). \tag{14.23}$$

Theorem 3. *The manufacturer's optimal wholesale price* w^* *is*

$$w^* = \frac{c_{\text{T}} + k c_{\text{M}}}{k - 1}. \tag{14.24}$$

Proof. Taking the first derivative with respect to w, we have

$$\frac{d\pi_{\text{m}}(w)}{dw} = y_0 z_0 \left((1 - F(z_0)) K_1 \right)^k (w + c_{\text{T}})^{-k-1} [(1-k)w + c_{\text{T}} + k c_{\text{M}}].$$

The sign of $\mathrm{d}\pi_{\text{m}}(w)/\mathrm{d}w$ is the same as that of $(1-k)w + c_{\text{T}} + kc_{\text{M}}$. Apparently, $\pi_{\text{m}}(w)$ increases before w reaches $(c_{\text{T}} + kc_{\text{M}})/(k-1)$ and starts to decrease after this point because $k > 1$. That is, $\pi_{\text{m}}(w)$ is unimodal in w and therefore the optimal wholesale price that maximizes $\pi_{\text{m}}(w)$ is $(c_{\text{T}} + kc_{\text{M}})/(k-1)$. This completes the proof. □

Remark 3. (1) w^* is greater than c_{M} because $k > 1$, which guarantees that the manufacturer always earns a positive profit.

(2) w^* is decreasing in k, which implies that the manufacturer should decrease his wholesale price if the market demand is more price sensitive.

(3) We can derive the profits π_{m}^* and π_{d}^* achievable by the manufacturer and the distributor, respectively, and compare their ratio

$$\beta := \frac{\pi_{\text{d}}^*}{\pi_{\text{m}}^*} = k \left(\frac{k}{k-1} \cdot \frac{K_0}{K_1} - 1 \right). \tag{14.25}$$

Because $K_0 \geq K_1$, we have $\beta \geq k(\frac{k}{k-1} - 1) = \frac{k}{k-1} > 1$. Therefore, without coordination, the manufacturer's profit might be much lower than that of the distributor.

(4) We can see that β increases in ρ. This means that the manufacturer's share of the profit decreases as the distributor becomes more conservative toward the transportation risk. Whether the manufacturer should coordinate with the distributor to share his transportation risk, so that everyone will be better off? This is an important question facing the manufacturer, which we will further investigate in Sect. 14.6.

14.5 Optimal Decisions in the Centralized System

The uncoordinated decisions derived in Sect. 14.4 may not be optimal from the point of view of the entire supply chain. We will investigate, in this section, the optimal decisions under a joint objective in the centralized system.

In the centralized system the manufacturer and the distributor seek to achieve the maximization of their objective (14.5). First, note that the optimal selling price of the distributor depends on the transportation time t and the order quantity q_c (see Theorem 1). Thus, given order quantity q_c of the distributor and transportation time t, the optimal selling price $p_c^*(q_c, t)$ can be obtained by Theorem 1.

On the other hand, we can show that (14.5) can be written as a function of q_c:

$$U_c(q_c) = \Pi_c(q_c) - \rho E_t\{[\Pi_c(q_c) - \Pi_c(q_c|t)]^+\}, \tag{14.26}$$

where $\Pi_c(q_c) = E_t\{\Pi_c(q_c|t)\}$, and $\Pi_c(q_c|t)$ is the maximal expected profit provided that the transportation time is t:

$$\Pi_c(q_c|t) = E_\varepsilon\{p_c^*(q_c,t)\min(q_c m(t), D(p_c^*(q_c,t),t))\} - (c_M + c_T)q_c. \tag{14.27}$$

Note that the wholesale price w of the manufacturer disappears in (14.26) and (14.27). This is because in the centralized system, the manufacturer's price w becomes an internal parameter, which does not affect the profit of the centralized system at all. Thus, the optimal decisions to be taken by the centralized system are the order quantity and the selling price in the market. The optimal selling price p_c is given by Theorem 1, whereas the optimal order quantity q_c is to be determined to maximize (14.26). These results are summarized in the theorem below.

Theorem 4. *In the centralized system, the optimal decisions are as follows:*

- *Optimal order quantity:*

$$q_c^* = y_0 z_0 \left(\frac{1-F(z_0)}{c_T + c_M} K_1\right)^k. \tag{14.28}$$

- *Optimal selling price, where t is the arrival time of the product at the market and q_c is the order quantity:*

$$p_c^*(q_c,t) = \left(\frac{y_0 \theta(t) z_0}{q_c m(t)}\right)^{1/k}. \tag{14.29}$$

The corresponding expected profit and expected utility are

$$\Pi_c^* = y_0 z_0 (c_T + c_M)\left(\frac{k}{k-1}\cdot\frac{K_0}{K_1} - 1\right)\left(\frac{1-F(z_0)}{c_T + c_M} K_1\right)^k, \tag{14.30}$$

and

$$U_{\rm c}^* = \frac{y_0 z_0 (c_{\rm T}+c_{\rm M})}{k-1}\left(\frac{1-F(z_0)}{c_{\rm T}+c_{\rm M}}K_1\right)^k. \tag{14.31}$$

Proof. Substituting (14.29) into (14.27), we have

$$\Pi_{\rm c}(q_{\rm c}|t) = \frac{k}{k-1}\left(y_0\theta(t)\right)^{1/k}\left(q_{\rm c}m(t)\right)^{1-1/k} z_0^{1/k}[1-F(z_0)] - (c_{\rm M}+c_{\rm T})q_{\rm c}.$$

Therefore,

$$\Pi_{\rm c}(q_{\rm c}) = E_t\{\Pi_{\rm c}(q_{\rm c}|t)\} = -(c_{\rm T}+c_{\rm M})q_{\rm c} + \frac{k}{k-1} y_0^{1/k} q_{\rm c}^{1-1/k} z_0^{1/k}[1-F(z_0)]K_0,$$

and

$$U_{\rm c}(q_{\rm c}) = \frac{k}{k-1} y_0^{1/k} q_{\rm c}^{1-1/k} z_0^{1/k}[1-F(z_0)]K_1 - (c_{\rm T}+c_{\rm M})q_{\rm c}.$$

For $\rho \in [0, K_0/E_t\{S(t)^+\})$, $U_{\rm c}(q_{\rm c})$ is concave in $q_{\rm c}$; hence, the optimal order quantity $q_{\rm c}^*$ is determined by the first-order condition, from which (14.28) is obtained. By substituting (14.28) into $\Pi_{\rm c}(q_{\rm c})$ and $U_{\rm c}(q_{\rm c})$, we have (14.30) and (14.31), respectively. This completes the proof. □

Note that the optimal order quantity now depends on information on the market demand (y_0, k and the demand distribution as contained in the stocking factor (see Lemma 1)), but not on the wholesale price w of the manufacturer. The optimal selling price can still be determined by a close-form relationship with the transportation time t: Given any realization of the random transportation time t, the optimal selling price can be determined by (14.29). See also Remark 1.

We now analyze the effects of coordination. We do so by examining the impact on the joint expected profit. We first present the following theorem.

Theorem 5. *The optimal centralized order quantity is always larger than that in the decentralized system:*

$$\eta := q_{\rm c}^*/q^* = \left(\frac{k}{k-1}\right)^k > 1.$$

Proof. It follows by comparing (14.21) with (14.28). □

The above result states that the ratio $q_{\rm c}^*/q^*$ depends only on the price elasticity of the market demand. This is obtained under the assumption that all the information is transparent to both the manufacturer and the distributor. It is not difficult to see that $[k/(k-1)]^k$ is decreasing in $k \in (1,\infty)$. Therefore, the more sensitive the market demand to the selling price (i.e., the larger the k value), the closer the optimal order quantity in the centralized system to that in the decentralized order quantity. This is in sharp contrast to the result of Weng (1999) where information is not transparent when there is a lack of coordination between the manufacturer and the distributor.

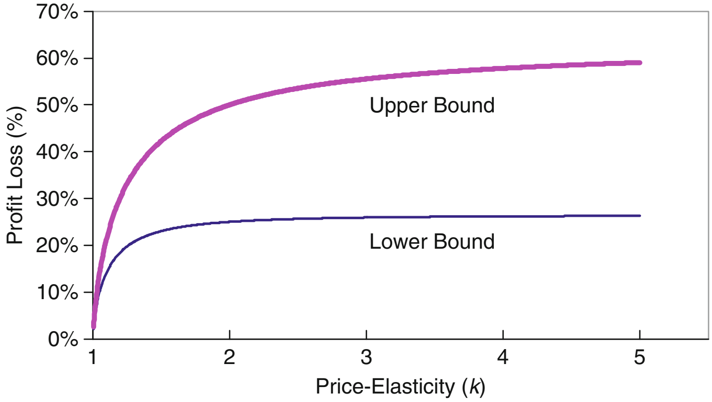


Fig. 14.2 The lower and upper bound of the profit loss

Next, we study the effects on the expected profit by coordination. Recall that the manufacturer and the distributor's expected profits in the absence of coordination are π_{m}^* and π_{d}^*, respectively. We define $\pi_{\mathrm{m}}^* + \pi_{\mathrm{d}}^*$ as the expected profit of the system in the absence of coordination. We are interested in the magnitude of the expected profit loss because of the lack of coordination between the manufacturer and the distributor. Denote $\varsigma := 1 - \frac{\pi_{\mathrm{m}}^*+\pi_{\mathrm{d}}^*}{\Pi_{\mathrm{c}}^*}$ as the expected profit loss due to lack of coordination. Hence, we have

$$\varsigma = 1 - \frac{1}{k-1}\left(\frac{k-1}{k}\right)^k \left[k + \left(\frac{k}{k-1}\cdot\frac{K_0}{K_1} - 1\right)^{-1}\right]. \tag{14.32}$$

It can be shown that when the decision makers are risk neutral (i.e., $\rho = 0$ and hence, $K_0 = K_1$), ς becomes $\varsigma^- = 1 - \frac{2k-1}{k-1}\left(\frac{k-1}{k}\right)^k$, which is increasing in k (as Fig. 14.2 shows). That is, the more sensitive the market demand to a change in price, the more profit loss incurred due to lack of coordination. The profit loss will become even larger if the decision makers are more risk averse toward the transportation loss because $K_0 \geq K_1$. Therefore, ς^- can be seen as the lower bound of the profit loss. It can be shown that as ρ increases (i.e., the decision makers are more risk averse), the profit loss also increases. We define $\varsigma^+ := 1 - \left(\frac{k-1}{k}\right)^{k-1}$ as the upper bound of the profit loss, which is obtained by letting $\rho \to K_0/E_t\{S(t)^+\}$. The lower bound and upper bound of the profit loss are plotted in Fig. 14.2.

We can see that the profit loss due to lack of coordination could be substantial, in particular, when the market demand is very price sensitive or the distributor is very risk averse. It follows from Theorem 5 that the distributor tends to order less in the decentralized system and consequently, the manufacturer would earn less when there is no coordination. As a result, the manufacturer would be much more interested in coordination with the distributor. How to motivate the distributor to order up to the quantity q_c^* required for the centralized system? This is the key for coordination. We investigate this topic in the next section.

14.6 An Incentive Scheme to Induce Coordination

Coordination between the distributor and the manufacturer should result in the scenario that the optimal decisions of the centralized system (Theorem 4) are adopted so that a higher level of welfare is achieved for all parties. For any transportation time t and order quantity q_c, there should be no problem for the distributor to set the selling price as specified by the optimal one (14.29). The problem is whether the distributor is willing to order up to the quantity q_c^* required for the centralized system. Due to risks involved in long transportation and random market demand, the distributor tends to order less, as our analysis has indicated in the previous section. The manufacturer, on the other hand, should more welcome coordination because in the decentralized system he would earn less since the distributor orders less. In other words, the manufacturer is the one that directly benefits from coordination. Whether the coordination can be achieved? The key is, therefore, whether the manufacturer can motivate the distributor to order up to the quantity q_c^* by offering the latter the appropriate incentive.

To motivate the distributor to order more, the manufacturer may offer to share a certain portion of the distributor's risk, such as the transportation risk. A direct mechanism one might think of is the use of a "flexible" wholesale price. That is, the manufacturer sets a basic wholesale price w_0 and offers the distributor a compensation when a transportation delay actually occurs. The compensation rate may depend on the realized transportation time t; i.e., the longer the t, the more compensation to be given to the distributor to offset his loss due to the deterioration and decay of the product during transportation. Technically, the manufacturer may set his wholesale price as a decreasing function of t in this scheme; hence, we call it the "flexible wholesale price" mechanism and denote the wholesale price as $w(t)$. An essential assumption is that the wholesale price should not be less than the manufacturer's cost, i.e., $w(t) \geq c_M$ for $\forall t$.

Although the flexible wholesale price above seems simple and reasonable, it is, however, unfortunate that it cannot make the supply chain coordinated in our problem involving price-sensitive demand and uncertain transportation time. This can be shown as follows, by using a counterexample with $\rho = 0$: It follows from Theorem 2 that the distributor's optimal order quantity $\hat{q}$ is given as

$$\hat{q} = y_0 z_0 \left(\frac{1 - F(z_0)}{\bar{w} + c_T} K_0 \right)^k, \tag{14.33}$$

where $\bar{w}$ is the mean of $w(t)$. This means that the distributor's optimal order quantity depends only on the average level of the wholesale policy. To achieve $\hat{q} = q_c^*$, the relation $\bar{w} = c_{\mathrm{M}}$ must hold; that is, the average wholesale price should be set equal to the unit manufacturing cost. It is apparent that the expected profit of the manufacturer will equal to zero, which would not be acceptable by the manufacturer. In such a case, coordination is not feasible.

We now propose a new incentive scheme. Our scheme consists of two parts:

(1) A price-discount sharing contract, under which the manufacturer shares a certain portion of the price discount that the distributor has to mark due to deterioration and decay of the product;
(2) A compensation contract (similar to the traditional buyback arrangement), under which the manufacturer compensates the distributor a certain amount for any unsold unit of the product.

Specifically, our price-discount sharing contract suggests

$$w(p,t) = c_{\mathrm{M}} + \alpha[m(t)p - c_{\mathrm{T}} - c_{\mathrm{M}}], \tag{14.34}$$

where $w(p,t)$ is the wholesale price of the manufacturer, which depends on the transportation time and the actual selling price of the distributor, and α is a constant taking value in $(0,1)$, which we will discuss later in this section.

We can show that (14.34) is equivalent to the following

$$w_0(t) - w(p,t) = \alpha m(t)(p_0 - p), \tag{14.35}$$

where $w_0(t)$ is the base or gross wholesale price equal to $\alpha[m(t)p_0 - c_{\mathrm{T}}] + (1 - \alpha)c_{\mathrm{M}}$, and p_0 is the "list price" of the product in the market. The relationship (14.34) indicates that the manufacturer should offer a wholesale price discount that is dependent upon the distributor's price-discount in the market as well as $m(t)$—the market quantity of the product after transportation.

This incentive scheme can be regarded as an extension of the price-discount sharing (PDS) scheme of Lal et al. (1996) and Bernstein and Federgruen (2005). Our scheme depends, however, on the transportation time of the product, which we call a time-dependent price-discount sharing (TDPDS) scheme.

Moreover, for any unit of the product that the distributor has failed to sell in the market, the manufacturer compensates the distributor an amount v as follows:

$$v = \alpha p. \tag{14.36}$$

Theorem 6. *TDPDS (14.34) together with the compensation contract (14.36) will induce the distributor to order the quantity q_{c}^* for any* $0 < \alpha < 1$.

Proof. Under the TDPDS scheme, the distributor's profit is as follows:

$$\begin{aligned}\pi_{\mathrm{d}}(p|q,t) &= p \cdot E_\varepsilon\{\min\{D(p,t), qm(t)\}\} - (w + c_{\mathrm{T}})q + vE_\varepsilon\{[qm(t) - D(p,t)]^+\} \\ &= pqm(t) - pE_\varepsilon\{[qm(t) - D(p,t)]^+\} - (w + c_{\mathrm{T}})q \\ &\quad + vE_\varepsilon\{[qm(t) - D(p,t)]^+\}\end{aligned}$$

$$= [m(t)p - w - c_T]q - (p - v)E_\varepsilon\{[qm(t) - D(p,t)]^+\}$$
$$= (1-\alpha)[m(t)p - c_M - c_T]q - (1-\alpha)pE_\varepsilon\{[qm(t) - D(p,t)]^+\}$$
$$= (1-\alpha)\Pi_c(p|q,t), \tag{14.37}$$

where the second equality is due to $\min(a,b) = b - (b-a)^+$. Thus, any optimal solution $p_c^*(t,q)$ and q_c^* that maximizes U_c optimizes U_d as well. That is, if the manufacturer applies the scheme (14.34) and (14.36), while seeking to maximize his expected utility, the distributor will order up to q_c^*. Therefore, the optimal welfare of the entire supply chain is achieved. □

Remark 4. (1) Suppose that the total profit achieved by the centralized system is Π_c^*. We can show that the manufacturer and the distributor's shares of the profit will be $\alpha\Pi_c^*$ and $(1-\alpha)\Pi_c^*$, respectively (cf. (14.37)).

(2) To ensure that each party's profit with coordination is not less than that before, we can derive a lower bound α^- and upper bound α^+, respectively:

$$\alpha^- = \pi_m^*/\Pi_c^* \text{ and } \alpha^+ = 1 - \pi_d^*/\Pi_c^*. \tag{14.38}$$

(3) The final choice of $\alpha \in [\alpha^-, \alpha^+]$ depends on the chain members' bargaining powers, but all such choices ensure that both parties will be better off by coordinating with each other.

14.7 Computational Studies

We have conducted a series of computational experiments to evaluate the effects of the following parameters on the optimal decisions in order to uncover certain managerial insights that are not obvious in the theoretical results:

- The fresh duration τ
- The downside penalty factor ρ
- The price elasticity k

The experiments have been designed as follows. The unit cost parameters are normalized as $c_M = 1$ and $c_T = 0.5$. The transportation time T is assumed to follow a uniform distribution in $[a,b]$ with $a = 1$ and $b = 5$. The product decays at a constant exponential rate $\gamma = 0.1$ beyond the fresh duration, i.e., $m(t) = \exp(-\gamma(t-\tau))$ for $t \in (\tau, b]$, and the value drop is also exponential with rate $\delta = 0.2$, i.e., $\theta(t) = \exp(-\delta(t-\tau))$ for $t \in (\tau, b]$. In the demand function, $y_0 = 100$, the price elasticity $k = 2$, and the random factor ε follows a normal distribution with mean 1 and variance $\sigma = 0.2$. The downside penalty factor $\rho = 1$.

In the first group of experiments, we changed the value of fresh durations τ from 2 to 5. For each τ, we derived the optimal decisions for both decentralized

Table 14.1 Optimal decisions under different fresh durations

	Decentralized				Centralized			
τ	w^*	q^*	π_{m}^*	π_{d}^*	q_{c}^*	Π_{c}^*	$\beta = \frac{\pi_{\mathrm{d}}^*}{\pi_{\mathrm{m}}^*}$	$\varsigma = 1 - \frac{\pi_{\mathrm{m}}^* + \pi_{\mathrm{d}}^*}{\Pi_{\mathrm{c}}^*}$
2.00	2.50	1.47	2.21	5.03	5.89	10.07	2.28	28.05%
2.25	2.50	1.57	2.35	5.30	6.27	10.60	2.25	27.82%
2.50	2.50	1.66	2.49	5.55	6.64	11.10	2.23	27.56%
2.75	2.50	1.75	2.63	5.79	7.02	11.58	2.20	27.28%
3.00	2.50	1.85	2.77	6.01	7.39	12.02	2.17	26.97%
3.25	2.50	1.94	2.90	6.21	7.74	12.43	2.14	26.64%
3.50	2.50	2.02	3.03	6.40	8.08	12.79	2.11	26.32%
3.75	2.50	2.10	3.15	6.55	8.39	13.11	2.08	25.99%
4.00	2.50	2.17	3.25	6.69	8.67	13.37	2.06	25.69%
4.25	2.50	2.23	3.34	6.79	8.90	13.59	2.03	25.42%
4.50	2.50	2.27	3.41	6.87	9.09	13.74	2.02	25.20%
4.75	2.50	2.30	3.45	6.92	9.20	13.84	2.00	25.06%
5.00	2.50	2.31	3.46	6.93	9.24	13.87	2.00	25.03%

and centralized distribution systems and the corresponding performance measures. The results are summarized in Table 14.1.

Observation I:

- Note that τ is an index on the perishability of the product. As τ decreases, the quantity and quality losses of the product increases.
- Table 14.1 shows that if the product is more perishable, the distributor's share of profit in the decentralized scenario increases. This can be explained as follows: When the product is more perishable, the quantity ordered by the risk-averse distributor becomes smaller, which in turn results in a decrease in the manufacturer's profit. However, the distributor could partly mitigate his loss through properly adjusting the selling price. As a result, the ratio of the distributor's profit over the manufacturer's profit increases when the product becomes more perishable.
- The profit loss when there is no coordination, ς, strictly decreases in τ. This implies that cooperation between the manufacturer and the distributor is especially profitable when the product is highly perishable.

In the second group of experiments, we changed the value of downside penalty factors ρ from 0.5 to 5. Again, for each ρ, we derived the optimal decisions for both decentralized and centralized distribution systems and the corresponding performance measures. The results are summarized in Table 14.2.

Table 14.2 Optimal decisions under different downside penalty factors

	Decentralized				Centralized			
ρ	w^*	q^*	π_{m}^*	π_{d}^*	q_{c}^*	Π_{c}^*	$\beta = \frac{\pi_{\mathrm{d}}^*}{\pi_{\mathrm{m}}^*}$	$\varsigma = 1 - \frac{\pi_{\mathrm{m}}^*+\pi_{\mathrm{d}}^*}{\Pi_{\mathrm{c}}^*}$
0.50	2.50	1.58	2.37	5.05	6.31	10.10	2.13	26.57%
1.00	2.50	1.47	2.21	5.03	5.89	10.07	2.28	28.05%
1.50	2.50	1.37	2.06	5.01	5.49	10.02	2.43	29.44%
2.00	2.50	1.28	1.91	4.97	5.10	9.94	2.60	30.75%
2.50	2.50	1.18	1.77	4.92	4.73	9.85	2.78	31.98%
3.00	2.50	1.09	1.64	4.86	4.37	9.73	2.97	33.15%
3.50	2.50	1.01	1.51	4.80	4.03	9.59	3.18	34.26%
4.00	2.50	0.92	1.39	4.72	3.69	9.43	3.40	35.31%
4.50	2.50	0.84	1.27	4.62	3.38	9.25	3.65	36.31%
5.00	2.50	0.77	1.15	4.52	3.07	9.05	3.92	37.25%

Table 14.3 Optimal decisions under different price elasticities

	Decentralized				Centralized				
k	w^*	q^*	π_{m}^*	π_{d}^*	q_{c}^*	Π_{c}^*	$\beta = \frac{\pi_{\mathrm{d}}^*}{\pi_{\mathrm{m}}^*}$	$\eta = \frac{q_{\mathrm{c}}^*}{q^*}$	$\varsigma = 1 - \frac{\pi_{\mathrm{m}}^*+\pi_{\mathrm{d}}^*}{\Pi_{\mathrm{c}}^*}$
1.25	7.00	0.72	4.34	24.00	5.41	35.88	5.53	7.48	21.02%
1.50	4.00	1.21	3.63	12.15	6.28	21.04	3.35	5.20	25.02%
1.75	3.00	1.46	2.91	7.67	6.42	14.48	2.63	4.41	26.90%
2.00	2.50	1.47	2.21	5.03	5.89	10.07	2.28	4.00	28.05%
2.25	2.20	1.44	1.73	3.57	5.39	7.45	2.07	3.75	28.87%
2.50	2.00	1.36	1.36	2.63	4.88	5.66	1.94	3.59	29.52%
2.75	1.86	1.25	1.07	1.97	4.33	4.35	1.84	3.47	30.06%
3.00	1.75	1.11	0.83	1.48	3.75	3.33	1.78	3.37	30.54%

Observation II:

- When ρ increases from 0.50 to 5.00, the distributor's expected profit has a decrease of only 10.50%, whereas the manufacturer has a profit decrease of 51.27%.
- Therefore, the manufacturer should cooperate with the distributor, especially when the distributor is quite risk averse.
- The profit loss because of noncoordination, ς, is also increasing in ρ. This means the supply chain members should coordinate when they are risk averse.

The last group of experiments focus on the price elasticity k and the results are summarized in Table 14.3.

Observation III:

- As the market demand becomes more sensitive to the selling price, the manufacturer's optimal wholesale price strictly decreases in the absence of coordination.

- The profits for both the manufacturer and the distributor *do* strictly decrease in k. Hence, both of them should prefer a less price-sensitive market demand.
- The profit loss due to noncoordination is increasing in k. Therefore, if the market is very sensitive to the price, the coordination of the manufacturer and the distributor is more beneficial.

14.8 Concluding Remarks

Supply chains involving long distance transportation of fresh products have become increasingly common in international as well as domestic markets. The requirement of long distance transportation and the perishability of fresh products make the decisions faced by the members in such a supply chain quite different from the traditional pricing and ordering decisions. A key issue that the decision makers have to take into consideration is the possible quantity decrease and quality decline of the product during long distance transportation. Another important issue that has to be addressed is the dependence of the market demand upon the level of the freshness of the product when it reaches the market. We have developed, in this paper, a model to address the decision concerns in such supply chains. We have derived the optimal decisions of the manufacturer and the distributor, in both the decentralized system and the centralized system. We have further developed a new incentive scheme by taking into account the specific features of the fresh-product supply chains, to facilitate coordination between the manufacturer and distributor. Computational studies have also been conducted to examine the effects of those critical parameters to the optimal decisions in different situations.

In our model the deterioration function $\theta(t)$ and the decay (obsolescence) function $m(t)$ take general forms, which can therefore be used to model the nature of different products in reality. A very useful characterization of the optimal pricing decision that we have obtained is its close-loop dependence on the realization of the transportation time, which enables the price to be set according to the actual level of freshness and the effective supply of the product when it reaches the market. The dependence of the optimal order quantity of the distributor and the wholesale price of the manufacturer on the information and data of the problem, such as the distribution of the transportation time, decay, and deterioration functions, and the distribution of the market demand, also enable important managerial insights to be analyzed and revealed.

A general conclusion derived from our study is that the coordination between the manufacturer and the distributor is an important strategy to be considered and enhanced; this is especially so when the following three scenarios prevail:

- The product has a very short fresh duration;
- The supply chain members are very conservative toward the transportation risk; and
- The market demand is very sensitive to the distributor's selling price.

The inventive scheme that we have proposed ensures that both parties will all be better off by coordination. The distribution of the extra benefit generated by coordination to each party depends, however, on the actual bargaining power of each party (as reflected by the parameter α in the incentive scheme). The incentive scheme can serve as the basis of contracts between the two parties.

The investigation of fresh product supply chains involving uncertain transportation is relative new line of research. We have considered the situation where the transaction between the manufacturer and the distributor is based on an FOB basis. In reality, there are many different business models between the upstream manufacturer and the downstream distributor. Another common model in export business is the so called "CIF" (cost insurance and freight), in which the manufacturer bears the transportation cost and risk. Apparently the uncoordinated and coordinated decisions and schemes in CIF transactions will be different. This is a topic we are currently investigating. It is also possible that a third-party logistics provider is responsible for the transportation, whose participation into the supply chain will impose new issues on the decisions, strategies, and coordination of the supply chain members. This is an interesting problem for further research. Another interesting topic is to consider the situation in which the product is to be sorted into different grades based on their levels of freshness when reaching the market. Different prices will then be determined for different grades of the product. This problem seems to be much more difficult, due to the interactions of the market demands for different grades of the product. Other topics for further research include multiple distributors, multiple products, etc. We expect that the framework of the model and the related results we have established in the current paper can serve as a basis for these further studies.

Acknowledgements We gratefully acknowledge the support of (i) Research Grants Council of Hong Kong, GRF No. CUHK410509 and Direct Grant No. 2050415, for X. Q. Cai; (ii) National Natural Science Foundation of China, Research Fund No. 70890082 and 70621061, and the MOE, P.R. China, through the Project of the Key Research Institute of Humanities and Social Sciences in Universities, under Grant No. 08JJD630001, for J. Chen; (iii) National Natural Science Foundation of China, Research Fund No. 70601017 and 71071083 for Y. B. Xiao, and (iv) National Natural Science Foundation of China, Research Fund No. 70801035 and 71171105, for X. L. Xu.

References

Agrawal, V., & Seshadri, S. (2000). Risk intermediation in supply chains. *IIE Transactions*, *32*(9), 819–831.

Anupindi, R., & Bassok, Y. (1999). Supply contracts with quantity commitments and stochastic demand. In S. Tayur, M. Magazine, & R. Ganeshan (Eds.), *Quantitative models for supply chain management* (pp. 197–232). Boston: Kluwer.

Barlow, R. E., & Proschan, F. (1965). *Mathematical theory of reliability*. New York: Wiley.

Bernstein, F., & Federgruen, A. (2005). Decentralized supply chains with competing retailers under demand uncertainty. *Management Science*, *51*(1), 18–29.

Bookbinder, J. H., & Cakanyildirim, M. (1999). Random lead times and expedited orders in (Q, r) inventory systems. *European Journal of Operational Research*, *115*, 300–313.

Cachon, G. P. (2003). Supply chain coordination with contracts. Chapter 6. In S. Graves & T. Kok, *Handbooks in operations research and managment science: Supply chain management*. New York: North-Holland publishers.

Cachon, G. P., & Lariviere, M. A. (2005). Supply chain coordination with revenue-sharing contracts: strengths and limitations. *Management Science*, *1*, 30–44.

Cai, X., Chen, J., Xiao, Y., & Xu, X. (2010). Optimization and coordination of fresh product supply chains with freshness-keeping effort. *Production and Operations Management*, *19*, 261–278.

Chen, F. (2003). Information sharing and supply chain coordination. In S. Graves & T. Kok, *Chapter 7, Handbooks in operations research and managment science: Supply chain management*. New York: North-Holland publishers.

Eppen, G. D., & Iyer A. V. (1997). Backup agreements in fashion buying—the value of upstream flexibility. *Management Science*, *43*(11), 1469–1484.

Eppen, G. D., & Martin, R. K. (1988). Determining safety stock in the presence of stochastic lead time and demand. *Management Science*, *34*(11), 1380–1390.

Feng, Y. Y., & Xiao, B. C. (2008). A combined risk and revenue analysis for managing perishable products. *Operations Research*, *56*(5), 1305–1311.

Goyal, S. K, & Giri, B. C. (2001). Recent trends in modeling of deteriorating inventory. *European Journal of Operational Research*, *134*, 1–16.

Hagen, J. W., Minami, D., Mason, B., & Dunton, W. (1999). California's produce trucking industry: characteristics and important issues. California Agricultural Technology Institute, http://cab.cati.csufresno.edu/research_publications/99/990301/.

Karaesmen, I. Z., Scheller-Wolf, A., & Deniz, B. (2011). Managing perishable and aging inventories: review and future research directions. In K. G. Kempf, P. Keskinocak, & R. Uzsoy (Eds.), *Planning production and inventories in the extended enterprise* (pp. 393–436). USA: Springer.

Kasmire, R. F. (1999). Fresh produce and perishability, RBCS Handbook, http://www.thepacker.com/rbcs/handbookarticles/properis.htm.

Krishan, H., Kapuscinski, R., & Butz, D. (2004). Coordinating contracts for decentralized supply chains with retailer promotional effort. *Management Science*, *50*(1), 48–63.

Lal, R., Little, J. D. C., & Villas-Boas, J. M. (1996). A theory of forward buying, merchandising, and trade deals. *Marketing Science*, *15*(1), 21–37.

Lodree Jr., E. J., & Uzochukwu, B. M. (2008). Production planning for a deteriorating item with stochastic demand and consumer choice. *International Journal of Production Economics*, *116*, 219–232.

Lau, H. S., & Zhao, L. G. (1993). Optimal ordering policies with two suppliers when lead times and demands are all stochastic. *European Journal of Operational Research*, *68*, 120–133.

Nahmias, S. (1982). Perishable inventory theory: a review. *Operations Research*, *30*(4), 680–708.

Nawrocki, D. (1999). A brief history of downside risk measures. *Journal of Investing*, *8*(3), 9–26.

Organisation for Economic Co-operation and Development (2005). *OECD in Figure-2005 edition*, http://www.oecd.org/document/.

Parlar, M., & Wang, Q. (1994). Discounting decisions in a supplier-buyer relationship under a linear buyer's demand. *IIE Transactions*, *26*, 34–41.

Pasternack, B. (1985). Optimal pricing and returns for perishable commodities. *Marketing Science*, *4*(2), 166–176.

Petruzzi, N. C., & Dada, M. (1999). Pricing and the newsvendor problem: a review. *Operations Research*, *47*(2), 183–194.

Plambeck, E., & Zenios, S. (2000). Performance-based incentives in a dynamic principal-agent model. *Manufacturing and Service Operations Management*, *2*, 240–263.

Raafat, F. (1991). Survey of literature on continuously deteriorating inventory models. *Journal of the Operational Research Society*, *42*, 27–37.

Rajan, A., Steinberg, R., & Steinberg, R. (1992). Dynamic pricing and ordering decisions by a monopolist. *Management Science*, *38*(2), 240–262.

Ray, S., Li, S. L., & Song, Y. Y. (2005). Tailored supply chain decision making under price-sensitive stochastic demand and delivery uncertainty. *Management Science*, *51*(12), 1873–1891.

Sanchez, J., & Kipe, S. (2004). U.S. Horticulture Exports to Asia 2004, Foreign Agricultural Service, Horticultural and Tropical Products Division, United States Department of Agriculture.

Shen, D., Lai, K. K., Leung, S. C. H., & Liang, L. (2010). Modelling and analysis of inventory replenishment for perishable agricultural products with buyer-seller collaboration. *International Journal of Systems Science*, *10*, 1–11.

Song, X. P., Cai, X., & Chen, J. (2005). Studies on interaction and coordination in supply chains with perishable products: a review. In C. K. Chan & H. W. J. Lee (Eds.), *Successful strategies in supply chain management, Chapter X* (pp. 222–248), Idea Group PA: Inc.

Taylor, T. (2000). Coordination under Channel Rebates with Sales Effort Effect. Working Paper, Stanford University, Stanford.

Tsay, A. (1999). The quantity flexibility contract and supplier-customer incentives. *Management Science*, *45*(10), 1339–1358.

Tsay, A., S. Nahmias, & Agrawal, N. (1999). Modeling supply chain contracts: a review. In S. Tayur, R. Ganeshan, & M. Magazine (Eds.), *Chapter 10, Quantitative Models for Supply Chain Management*. Boston: Kluwer Academic.

Wang, Q. N. (2001). Coordinating independent buyers in a distribution system to increase a supplier's profits. *Manufacturing & Service Operation Management*, *3*, 337–348.

Wang, Y. (2004). Joint Pricing-production Decisions in Supply Chains of Complementary Products with Uncertain Demand. Working Paper, Weatherhead School of Management, Case Western Reserve University.

Wang, Y., Jiang, L., & Shen, Z. (2004). Channel performance Under consignment contract with revenue sharing. *Management Science*, *50*(1), 34–47.

Weng, Z. K. (1995). Channel coordination and quantity discounts. *Management Science*, *41*, 1509–1522.

Weng, Z. K. (1999). The power of coordinated decisions for short-life-cycle products in a manufacturing and distribution supply chain. *IIE Transactions*, *31*, 1037–1049.

Whitin, T. M. (1957). Theory of inventory management. NJ: Princeton University Press.

Chapter 15
Profit Target Setting for Multiple Divisions: A Newsvendor Perspective

Chunming (Victor) Shi and Lan Guo

Abstract Managers and firms often engage in decision making based on certain profit targets. Consequently, they may adopt the objective of maximizing the profit probability, namely, the probability of achieving those profit targets. However, there has been limited research on modeling profit target setting. In this chapter, we study analytic target setting under a common business scenario where a firm owns multiple divisions. The firm sets a profit target for each division, which then decides on production level and selling price to maximize the profit probability. We obtain the divisions' optimal profit targets in closed forms when the firm's objective is to maximize its expected profit. When the firm's own objective is also to maximize profit probability, the problem of profit target setting is more complicated. To gain more managerial insights, we focus on two specific cases. In the first case of fair target setting, we show that for most reasonable customer demand distributions, if a division has a relatively high (low) production cost, its assigned profit target decreases (increases) in its price elasticity. In the second case, if the firm is in control of two identical divisions, each division's optimal profit target is just half of the firm's profit target when the price elasticity is two or more, regardless of production cost and demand distribution. We hope that the managerial insights from this chapter help practitioners who are involved with target setting and target attainment.

Keywords Newsvendor • Pricing • Risk aversion • Target setting

C. (Victor) Shi (✉) • L. Guo
Operations and Decision Sciences, School of Business and Economics, Wilfrid Laurier University, 75 University Avenue West, Waterloo, Ontario, Canada, N2L 3C5
e-mail: cshi@wlu.ca; LAGUO@WLU.CA

T.-M. Choi (ed.), *Handbook of Newsvendor Problems*, International Series in Operations Research & Management Science 176, DOI 10.1007/978-1-4614-3600-3_15,

15.1 Introduction

A great deal of research in Operations Management assumes the objective of maximizing expected profit or, sometimes equivalently, minimizing expected cost. However, in business practice, employees, managers, and firms often engage in decision making based on profit targets (Bordley and Kirkwood 2004; Abbas and Matheson 2005; Abbas et al. 2009). That is, they either are assigned profit targets by some external forces such as the corporate headquarters and analysts or set profit targets themselves through participative budgeting process, and they are rewarded or penalized based on whether they meet those targets or miss them (Jensen 2003). As a result, they may adopt the objective of maximizing profit probability, namely, the probability of reaching those profit targets.

The difference between the objective of maximizing expected profit and that of maximizing profit probability is by no means trivial. Jensen (2003) provides examples of how the latter objective drives managerial behaviors that diverge from those driven by the former objective. In one case, a manager knows the target is far from reach and even if he tries his best, he is unlikely to reach the profit target. In the other case, a manager knows that he can easily beat the target and increasing profit will not bring him extra earnings. In both cases, the manager would not be motivated to maximize expected profit.

Assuming objective of maximizing profit probability has two advantages over assuming objective of maximizing expected profit. First, given a profit target exogenously set, the objective of maximizing profit probability assumes risk aversion by definition. To be more specific, it operationalizes risk aversion through a critical probability, specifically, the probability that the profit is no less than a certain threshold. Note that there are two closely related risk measures including Value-at-Risk and Conditional Value-at-Risk (Gan et al. 2005; Ozler et al. 2009). With those two measures, decision makers maximize expected profit while controlling for some critical probability. It is certainly worthwhile for future research to incorporating those more complicated measures when setting profits for divisions.

The second advantage is that under some situations, the objective of maximizing profit probability is more descriptive of how firms and managers make decisions. A big literature on public firms' earnings management behaviors (e.g., Burgstahler and Dichev 1997; Degeorge et al. 1999; Healy and Wahlen 1999; Dechow and Skinner 2000) indicates that meeting or beating various profit targets or thresholds (e.g., prior years' profit, analysts' forecasts, avoiding losses, and performance targets specified in executive compensation contracts) is the most important motive for firms to manipulate accounting numbers. Although such unlawful or unethical behavior is not carried out in most firms, it does indicate the extreme importance of meeting or beating profit targets at the firm level. One such example involves eBay. In the 4th quarter of 2004, eBay's reported profit of 33 cents per share missed the profit target of 34 cents, which was set by Wall Street analysts. Although the gap was only one cent, eBay's stock price fumbled by 12% right after the report (http://money.cnn.com/2005/01/19/news/fortune500/ebay/index.htm). For a more recent example, Swiss Life Holding, Switzerland's largest life insurer, missed its profit target of $1.6 billion for the year 2008. Consequently, its stock price fell 20% in Zurich trading (Giles 2008).

It has been well documented that the objective of maximizing profit probability is also common and important for divisional managers. As shown by Bouwens and Van Lent (2007), it is not uncommon for firms to evaluate their divisional managers based on their performance on profits. Paying managers based on their actual profits relative to a profit budget or target is also often seen in practice (Jensen 2003). In a recent paper, Brown and Tang (2006) interview six buyers from different firms and find that product profit and gross margin are ranked as the most important performance measures in their firms, and targets for profit and gross margin are often used. The authors use profit-target-based reward system to explain the irregularity they observe in their experimental setting, i.e., participants consistently select order quantities less than the newsvendor solution.

Limited research has been done on the objective of maximizing profit probability in the Operations Management literature. The earliest work includes Kabak and Schiff (1978) and Lau (1980). These two papers study a newsvendor with the objective of maximizing profit probability. Lau and Lau (1988) and Li et al. (1990) study the problem of a newsvendor selling two products with the objective of maximizing profit probability. These two studies focus on the special case where the customer demand for each product has a uniform distribution. Li et al. (1991) focus on the special case of exponential distribution. Parlar and Weng (2003) study the newsvendor model where the objective is to maximize the probability of achieving the expected profit, which is a function of order quantity. Recently, the study on the objective of maximizing profit probability is extended to the framework of supply chains. Shi and Chen (2007) study a basic supply chain where a single supplier sells to a newsvendor-type retailer and both adopt the objective of maximizing profit probability. Contrary to the result under the objective of maximizing expected profit, they show that a properly designed wholesale price contract can coordinate the supply chain.

It appears that the extant literature usually assumes that profit targets are exogenously set. However, it is of strategic importance to study profit target setting, i.e., how the values of profit targets are set. Too high a profit target provokes frustration and cynicism, whereas too low a profit target causes apathy and lacks motivational value. As has long been documented in the organizational behavior literature (for a review, see Locke 2001; Locke and Latham 1990, 2002), target or goal difficulty has a significantly positive effect on task performance until limits of ability are reached or individuals cease to be committed to the highly difficult target. As for how firms actually set their profit targets, Merchant and Manzoni's (1989) field study provides some interesting insights. They find that among firms with multiple divisions, 80% to 90% of the time they set their annual profit targets at achievable levels (vs. stretch goals). However, to best of our knowledge, little research has been done on modeling profit target setting.

There are three papers (Lau and Lau 1988; Li et al. 1990, 1991) that relate to analytic target setting indirectly. In these three papers, the authors study the two-product newsvendor problem where order quantity and profit target are related. However, their focus is on determining the optimal order quantities under the uniform and exponential demand distributions. Further, the authors assume that selling prices of the products are exogenously fixed. Finally, it is noted that there has been some research (see, e.g., Yang et al. 2009; Lovell and Pastor 1997) on target setting using the DEA, which is basically a nonparametric methodology based on deterministic linear programming. Consequently, DEA is not applicable to our research problem of profit target setting which involves stochastic customer demand.

In this chapter, we attempt to address the profit target setting issue for multiple divisions within a firm. To be more specific, we study a business scenario where a firm owns multiple autonomous divisions that have authority over operations and other decision making. Such a scenario is common in modern decentralized corporations to allow for rapid and flexible decision making (Aghion and Tirole 1997; Roberts 2004). When a firm sells a single product or service in different countries or continents, one division can be set for each country or continent. For example, Dell has Dell USA, Dell Germany, and Dell China as three of its divisions. When a firm sells different products within the same market, one division can be set for each product or product category. An example is Hewlett-Packard, which is organized into three divisions: the personal systems group, the imaging and printing group, and the technology solutions group (http://www.hp.com/hpinfo/abouthp/). In general, the roles of the firm include allocating resources (e.g., capital) across divisions and specifying outside vendors for economies of scale and/or quality assurance. It is also possible that the firm supplies some products directly to its divisions.

We model a single division as a newsvendor with the objective of maximizing profit probability. Furthermore, we extend the existing literature by considering a price-setting newsvendor. That is, given a profit target, each division decides on production level and selling price simultaneously to maximize the probability of achieving its profit target. We also examine how the firm should set profit targets for its divisions, all of which have the objective of maximizing profit probability. Our results are derived in two specific cases. In the first case of fair target setting, we follow prior studies (e.g., Bushman et al. 1995) and assume that the sum of all divisional profit targets equals the profit target for the firm. In the second case, we assume the firm only owns two divisions.

The rest of the chapter is organized as follows. In Sect. 15.2, we derive the optimal production and pricing decisions of a single division given its assigned profit target. We proceed on studying the problem of profit target setting from the firm's perspective when a firm owns multiple divisions, each of which has a profit target. In Sect. 15.3, we study the firm with the objective of maximizing expected profit. In Sect. 15.4, we study the firm with the objective of maximizing profit probability, where we focus on the special cases of fair target setting and two identical divisions. Finally, in Sect. 15.5, we summarize and discuss future research directions.

15.2 A Single Division Given a Profit Target

Before we study the problem of profit target setting for multiple divisions, we need to derive the optimal production level and selling price of a single division given a profit target. This is the focus of this section.

Suppose that a division has a unit production cost (or procurement cost) of c and a unit selling price of r. To simplify the presentation, both salvage value and loss-of-goodwill cost are assumed to be zero. Furthermore, instead of the traditional objective of maximizing expected profit, the division is assigned a profit target t and, hence, adopts the objective of maximizing profit probability. In other words, the division maximizes the probability of achieving the profit target t. For simplicity, the probability is called the profit probability.

It is worth noting that targets may not be always set to maximize profit. Under uncertain business environment (e.g., stochastic customer demand), only expected profit can be maximized. However, even if *expected* profit is maximized, it is likely that the *actual* profit is lower than the expected profit due to the variability in profit. This is particularly a concern because by definition, newsvendors make one-time and, hence, nonrepeatable decisions.

The market demand is random and is affected by selling price. In this research, the market demand is modeled as a multiplicative form:

$$D(r) = r^{-b}\varepsilon \tag{15.1}$$

where ε is a random variable taking positive values with cumulative distribution function $F(x)$ and probability density function $f(x)$. To avoid trivial situations, $F(x)$ is assumed to be increasing and differentiable. The price elasticity b represents the extent to which the customers are sensitive to price changes. To be more specific, if the price changes by 1%, the customer demand changes by b% in the opposite direction. Some examples of price elasticity for a product category include fresh green peas ($b = 2.8$), fresh tomatoes ($b = 4.6$), and Chevrolet automobiles ($b = 4.0$) (Gwartney 1976). In this chapter, it is assumed that each division is selling an individual product, which tends to have a greater price elasticity mainly due to the great availability of substitutes in the same product category (Gwartney 1976). Therefore, it is reasonable to limit ourselves to elastic goods or services ($b > 1$).

The multiplicative demand model is also called an isoelastic demand model or double-log-linear model. One limitation of such a model is that a price change results in a scale change, but *not* a location change, in the demand distribution. Despite this limitation, this demand model is the most frequently used demand specification among econometricians, market empiricists, and researchers in Operations Management. Monahan et al. (2004) summarize four reasons to explain the popularity of this multiplicative demand model:

1. It is consistent with consumer-utility-maximization theory; thus, it is a reasonable candidate for model building.

2. By explicitly accounting for the effects of price elasticity on demand, it has an unambiguous economic interpretation.
3. Its log-linearity is particularly amenable to empirical analysis because its parameters can be estimated using well-established linear regression techniques.
4. Perhaps most importantly, it typically provides a good statistical fit with available sales data.

Typical products whose demands may follow the multiplicative model include high fashion products or newly introduced products (Agrawal and Seshadri 2000).

The division's random profit function as a function of production level and selling price is given by:

$$\Pi(q,r) = (r-c)q - r(q-D(r))^{+}. \tag{15.2}$$

Now, the division is given a profit target t to achieve. For simplicity, the profit target t, once determined, is assumed to be independent of the market demand in this section. Of course, when a profit target is being set, a number of factors may be taken into account, including the market demand. This is the target-setting problem to be studied in later sections.

If selling price r is exogenous, the division's optimal production level and the maximal profit probability are given by (see, e.g., Kabak and Schiff 1978):

$$q(r) = \frac{t}{r-c} \tag{15.3}$$

$$P(r) = 1 - F_{D(r)}\left(\frac{t}{r-c}\right) = 1 - F\left(\frac{t}{(r-c)r^{-b}}\right), \tag{15.4}$$

where $F_{D(r)}(\cdot)$ denotes the cumulative distribution function of demand $D(r)$. To choose r to maximize the profit probability $P(r)$, it is equivalent to maximize $(r-c)r^{-b}$. It can be easily verified that the optimal selling price, denoted by $G(b,c)$, is given by:

$$G(b,c) = \frac{b}{b-1}c. \tag{15.5}$$

It can be seen from (15.5) that the greater the price elasticity, the lower the optimal selling price. More interestingly in the current context, the optimal selling price is independent of profit target t. This is because of the multiplicative demand model where pricing affects only the scale, but not the location, of the demand distribution. Finally, the optimal profit margin is:

$$\frac{G(b,c)-c}{G(b,c)} = \frac{1}{b}, \tag{15.6}$$

which is exactly the reciprocal of the price elasticity. Hence, a greater price elasticity always leads to a lower profit margin.

By substituting the optimal price $G(b,c)$ into (15.3) and (15.4), we have the optimal production level and the maximal profit probability as:

$$q^* = \frac{b-1}{c}t = H(b,c)t \tag{15.7}$$

$$\begin{aligned} P^* &= 1 - F(L(b,c)H(b,c)t) \\ &= 1 - F\left(\frac{c^{b-1}b^b}{(b-1)^{b-1}}t\right), \end{aligned} \tag{15.8}$$

where $H(b,c) = (b-1)/c$ and $L(b,c) = (bc/(b-1))^b$ are defined for notation simplicity.

It can be seen from (15.7) that the optimal production level q^*s increases with respect to the price elasticity. This is because a greater price elasticity leads to a lower profit margin. To achieve the same profit target, production level (and, hence, sales revenue) has to be larger. From (15.8), it can be seen that the higher the production cost and the higher the profit target, the smaller the maximal profit probability P^*. Both predictions make intuitive sense. To detail how price elasticity impacts the maximal profit probability, we have the following two propositions.

Proposition 1. *When production cost is relatively high, i.e., $c \geq (b-1)/b$, the term $L(b,c)H(b,c)$ increases with respect to the price elasticity b. Otherwise, the term $L(b,c)H(b,c)$ decreases with respect to the price elasticity b.*

Proof. It can be seen from (15.8) that:

$$L(b,c)H(b,c) = \frac{c^{b-1}b^b}{(b-1)^{b-1}}. \tag{15.9}$$

Differentiating $L(b,c)H(b,c)$ with respect to b, we have:

$$\frac{\partial L(b,c)H(b,c)}{\partial b} = \frac{c^{b-1}b^b}{(b-1)^{b-1}} \ln \frac{bc}{b-1}. \tag{15.10}$$

Therefore, if production cost is relatively high, i.e., $c \geq (b-1)/b$, we have $\ln(bc/(b-1)) \geq 0$ and, hence, the first derivative $\partial L(b,c)H(b,c)/\partial b \geq 0$. Otherwise, we have $\partial L(b,c)H(b,c)/\partial b < 0$. This concludes the proof. □

Proposition 2. *When production cost is relatively high, i.e., $c \geq (b-1)/b$, a greater price elasticity leads to a smaller maximal profit probability. When production cost is relatively low, i.e., $c < (b-1)/b$, a greater price elasticity leads to a larger maximal profit probability.*

Proof. It can be seen from (15.8) that the maximal profit probability P^* and the term $L(b,c)H(b,c)$ change with respect to the price elasticity b in *opposite* directions. Hence, Proposition 2 can be proved similarly to Proposition 1. □

Therefore, how maximal profit probability changes with respect to price elasticity depends on the value of production cost. The intuitions are as follows. To increase profit and, hence, profit probability, a division has two options. Option 1 is to increase selling price, which, however, reduces customer demand. Option 2 is to lower selling price, which increases customer demand. For a division with high price elasticity, Option 1 is unattractive because the benefit of price increase is more than offset by the cost of demand drop. Therefore, a division with high price elasticity should adopt Option 2. If the same division has relatively low production cost, the greater the price elasticity, the larger demand increase will result from the same level of price drop. Consequently, profit as well as profit probability will be higher. On the other hand, if the same division has a relatively high production cost, there is a limit in terms of price reduction because selling price has to be larger than production cost to have a positive profit. Therefore, with high price elasticity, the benefit of demand increase is insufficient to compensate for the cost of price drop. As a result, both profit and profit probability will be lower.

The discussion above implicitly assumes that the random variable ε, and, hence, the customer demand, is defined on $[0,+\infty]$, which implies any profit target t in theory could be achieved. However, we assume that only an achievable profit target will be assigned. We thus establish an upper bound on profit target where ε is defined on a limited interval. The result is given in the following proposition, which will be used in Sect. 15.4.

Proposition 3. *Suppose that the market demand is given by* $D(r) = r^{-b}\varepsilon$ *where the random variable* ε *is defined on* $[\alpha, \beta]$. *For a profit target* t *to be achievable at all, it is required that* $t < \frac{1}{c^{b-1}} \frac{(b-1)^{b-1}}{b^b} \beta$.

Proof. If the division sets selling price at r, then the maximal achievable profit is $(r-c)r^{-b}\beta$. This happens when the random variable ε is realized at its maximal possible value β. Furthermore, $(r-c)r^{-b}\beta$ is maximized at $G(b,c)$ (defined in (15.5)). Therefore, all feasible profit targets are less than $(G(b,c)-c)L^{-1}(b,c)\beta$, which is equivalent to $t < \frac{(b-1)^{b-1}}{c^{b-1}b^b}\beta$. This concludes the proof. □

15.3 A Firm with the Objective of Maximizing Expected Profit

Starting with this section, we will study the business scenario where a firm owns n divisions. Suppose each division's performance is evaluated based on whether or not it achieves a predetermined profit target. As a result, each division always adopts the objective of maximizing profit probability, i.e., the objective of maximizing the probability of achieving its profit target. On the other hand, the firm itself may adopt the objective of maximizing expected profit or the objective of maximizing profit probability, which will be addressed in this section and next section, respectively.

If the firm adopts the objective of maximizing expected profit, the firm is assumed to be risk neutral. This assumption of risk neutrality could be reasonable when the firm is able to diversify its risk through the n divisions, especially when n is relatively large.

Because of the objective of maximizing expected profit, maximizing expected profit of the firm is equivalent to maximizing the expected profit of each division assuming that intra-firm transactions and interdependence between divisions are negligible. This property holds true only because expected value is a linear operator. Since there are n divisions, the firm needs to assign n profit targets t_i, where subscript $i = 1 \cdots n$ denotes division i. Division i has a unit production cost c_i and a price elasticity b_i. Its customer demand is given by $D_i(r_i) = r_i^{-b_i}\varepsilon_i$, where the random variables ε_i (with cumulative distribution function $F_i(\cdot)$) is independent of ε_j, $j \neq i$.

It can be shown from (15.2) that the expected profit of division i is given by:

$$E\Pi_i^*(t_i) = t_i - G_i(b_i, c_i)\int_0^{H_i(b_i,c_i)t_i} F_i(L_i(b_i,c_i)x)\mathrm{d}x. \tag{15.11}$$

We have the following theorem on setting the optimal profit targets for the divisions.

Theorem 1. *If the firm is risk neutral, the divisions' optimal profit targets are given by:*

$$t_i^* = \frac{1}{c_i^{b_i-1}}\frac{(b_i-1)^{b_i-1}}{b_i^{-b_i}}F_i^{-1}\left(\frac{1}{b_i}\right), \quad i = 1\cdots n \tag{15.12}$$

Proof. Based on (15.11), we have the following derivatives:

$$\frac{\partial E\Pi_i^*(t_i)}{\partial t_i} = 1 - b_i F_i(L_i(b_i,c_i)H_i(b_i,c_i)t_i), \tag{15.13}$$

$$\frac{\partial^2 E\Pi_i^*(t_i)}{\partial t_i^2} = -b_i L_i(b_i,c_i)H_i(b_i,c_i)f_i(L_i(b_i,c_i)H_i(b_i,c_i)t_i). \tag{15.14}$$

Because $\partial^2 E\Pi_i^*(t_i)/\partial t_i^2 < 0$, $\Pi_i^*(t_i)$ is concave in t_i. By setting the first derivative to zero, we can obtain (15.12). This concludes the proof. □

Based on (15.12), an immediate observation is that a division with a higher production cost should be assigned a lower profit target. Moreover, it can be seen from Proposition 1 that when production cost is relatively high, $L_i(b_i,c_i)H_i(b_i,c_i)$ increases in price elasticity and, hence, a division with a greater price elasticity will be assigned a lower profit target. However, when production cost is relatively low, because both the denominator $L_i(b_i,c_i)H_i(b_i,c_i)$ and the nominator $F_i^{-1}(1/b_i)$ decrease in b, the optimal profit target may increase or decrease with price elasticity.

15.4 A Firm with the Objective of Maximizing Profit Probability

In this section, we assume the firm itself has a profit target T to achieve, and the firm adopts the objective of maximizing profit probability, i.e., the objective of maximizing the probability of achieving the target T. Its decisions are to assign profit target t_i to division i, $i = 1 \cdots n$.

Given a profit target t_i, division i chooses the optimal selling price as in (15.5) and the optimal production level as in (15.7). Substituting (15.5) and (15.7) into (15.2), we have the associated random profit of division i as:

$$\Pi_i(t_i) = t_i - G_i(b_i,c_i)(H_i(b_i,c_i)t_i - L_i^{-1}(b_i,c_i)\varepsilon_i)^+. \tag{15.15}$$

The random profit function of the firm is then given as:

$$\Pi(t_1,\cdots t_n) = \sum_{i=1}^n \left[t_i - G_i(b_i,c_i)(H_i(b_i,c_i)t_i - L_i^{-1}(b_i,c_i)\varepsilon_i)^+\right]. \tag{15.16}$$

It can be seen from (15.16) that the maximal achievable profit target for the firm is the sum of individual profit target, i.e., $\sum_{i=1}^n t_i$. This is because the function $(\cdot)^+$ only takes non-negative values. Therefore, to make sure the profit target is achievable for the firm, it is required that $\sum_{i=1}^n t_i \geq T$.

The firm's optimization problem is to select profit target t_i such that its own profit probability is maximized. Mathematically, it is formulated as follows:

$$\begin{aligned}\max_{t_i} P_T(t_1,\cdots t_n) &= \max_{t_i} P\{\Pi(t_1,\cdots t_n) \geq T\} \\ &= \max_{t_i} P\left\{\sum_{i=1}^n G_i(b_i,c_i)\left(H_i(b_i,c_i)t_i - L_i^{-1}(b_i,c_i)\varepsilon_i\right)^+ \leq \sum_{i=1}^n t_i - T\right\}.\end{aligned} \tag{15.17}$$

Unfortunately, this optimization problem in general does not have analytic solutions and may require the use of numerical simulation. To simplify the optimization problem and thus gain more managerial insights, we study two simplified cases, namely, the case of fair target setting and the case of two divisions in the following two sections, respectively. Further, for concision, we denote the functions $G_i(b_i,c_i)$, $H_i(b_i,c_i)$ and $L_i(b_i,c_i)$ by G_i, H_i, and L_i, respectively, when no confusion arises.

15.4.1 Fair Target Setting Case

In the case of fair target setting, the sum of the divisions' profit targets is equal to the profit target of the firm, i.e., $\sum_{i=1}^n t_i = T$ (see, e.g., Lau and Lau 1988). This is consistent with Bushman et al. (1995) who assumes that the sum of the multiple divisions' outputs is the firm's output.

Under the requirement of fair target setting, i.e., $\sum_{i=1}^n t_i = T$, the optimization problem (15.17) can be simplified to:

$$\max_{t_i} P\left\{\sum_{i=1}^n G_i\left(H_i t_i - L_i^{-1}\varepsilon_i\right)^+ \leq 0\right\}. \tag{15.18}$$

Because the function $(\cdot)^+$only takes nonnegative values and random variables ε_i's are independent of each other, the optimization problem (15.18) can be further simplified to:

$$\max_{t_i} P\left\{L_i^{-1}\varepsilon_i \geq H_i t_i,\ i=1,\cdots n\right\} = \max_{t_i} \bar{F}_1(L_1H_1t_1)\bar{F}_2(L_2H_2t_2)\cdots\bar{F}_n(L_nH_nt_n), \tag{15.19}$$

where $\bar{F}_i(x) = 1 - F_i(x)$. We further define $W_i(x) = f_i(x)/\bar{F}_i(x)$ as the failure rate (or the hazard rate) of random variable ε_i. A demand distribution exhibits the property of increasing failure rate if $\partial W_i(x)/\partial x > 0$. Examples of such distributions include most reasonable customer demand distributions such as uniform, exponential, normal, gamma, and Weibull distributions (Barlow and Proschan 1965; Lariviere 2006).

Theorem 2. *If each division's demand distribution has an increasing failure rate, under fair target setting, the optimal profit targets for the divisions can be solved from the following equations:*

$$W_1(L_1H_1t_1^*)L_1H_1 = \ldots = W_n(L_nH_nt_n^*)L_nH_n \text{ and } \textstyle\sum_{i=1}^n t_i^* = T. \tag{15.20}$$

Further, for a division with a relatively high (low) production cost, a greater price elasticity leads to a lower (higher) optimal profit target.

Proof. To maximize (15.19), it is equivalent to maximize

$$\textstyle\sum_{i=1}^n \ln\bar{F}_i(L_iH_it_i) \tag{15.21}$$

subject to the constraint $\sum_{i=1}^n t_i = T$. The Lagrangian function for this optimization problem is:

$$Z(t_1,\cdots t_n,\lambda) = \textstyle\sum_{i=1}^n \ln\bar{F}_i(L_iH_it_i) = -\lambda\left(\sum_{i=1}^n t_i - T\right). \tag{15.22}$$

The first-order derivatives are given by:

$$\frac{\partial Z(t_1,\cdots t_n,\lambda)}{\partial t_i} = -L_iH_i\frac{f_i(L_iH_it_i)}{\bar{F}_i(L_iH_it_i)} - \lambda = -L_iH_iW_i(L_iH_it_i) - \lambda, \quad i=1,\cdots n \tag{15.23}$$

$$\frac{\partial Z(t_1,\cdots t_n,\lambda)}{\partial \lambda} = \textstyle\sum_{i=1}^n t_i - T \tag{15.24}$$

The second-order derivatives with respect to t_i are given by:

$$\frac{\partial^2 Z(t_1,\cdots t_n,\lambda)}{\partial t_i^2} = -L_i^2H_i^2\frac{\partial W_i(L_iH_it_i)}{\partial t_i}. \tag{15.25}$$

Because the demand is modeled as $D_i(r_i) = r_i^{-b_i}\varepsilon_i$, the distribution of D_i has an increasing failure rate if and only if the distribution of ε_i has an increasing failure rate $W_i(x)$. Therefore, we have $\partial W_i(L_iH_it_i)/\partial t_i > 0$ and thus $\partial^2 Z(t_1,\cdots t_n,\lambda)/\partial t_i^2 < 0$.

Hence, $Z(t_1,\cdots t_n,\lambda)$ is concave in $(t_1,\cdots t_n)$. Setting the first-order derivatives (15.23) and (15.24) to zero, we can obtain (15.20).

Suppose division i has a relatively high production cost. It can be seen from Proposition 1 that the term $L_i(b,c)H_i(b,c)$ increases in the price elasticity b_i. Based on (15.20), the optimal profit target then decreases in b_i because $W_i(x)$ is an increasing function. Using similar arguments, we can show that the opposite is true for a division with a relatively low production cost. □

Example 1. Suppose that ε_i follows uniform distribution defined on interval $[0,\beta_i]$. So we have:

$$\bar{F}_i(x) = \frac{\beta_i - x}{\beta_i}, \quad 0 \le x \le \beta_i \text{ and } W_i(x) = \frac{1}{\beta_i - x}. \tag{15.26}$$

Based on (15.20), we have:

$$\frac{L_1H_1}{\beta_1 - L_1H_1t_1^*} = \frac{L_2H_2}{\beta_2 - L_2H_2t_2^*} = \ldots = \frac{L_nH_n}{\beta_n - L_nH_nt_n^*}, \tag{15.27}$$

which further gives:

$$t_j^* = t_i^* + \frac{\beta_j}{L_jH_j} - \frac{\beta_i}{L_iH_i}, \quad i,j = 1,\cdots n \text{ and } i \ne j. \tag{15.28}$$

Together with $\sum_{i=1}^{n} t_i^* = T$, we have the optimal profit target for each division:

$$t_i^* = \frac{T}{n} + \frac{1}{n}\sum\nolimits_{j=1,j\ne i}^{n}\left(\frac{\beta_i}{L_iH_i} - \frac{\beta_j}{L_jH_j}\right), \quad i = 1,\cdots n \tag{15.29}$$

Therefore, the optimal profit target consists of two components. The first component is the firm's profit target divided by the number of divisions. The second component shows how the optimal profit target for each division is further adjusted by its market size β_i and the term $L_i(b,c)H_i(b,c)$, which increases (decreases) with respect to the price elasticity b_i when production cost c_i is relatively low (high).

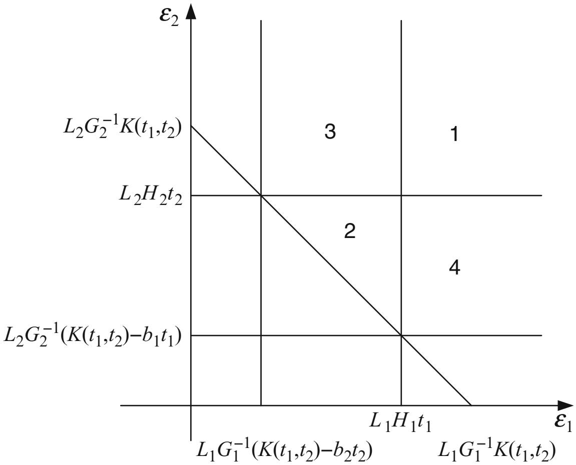
15.4.2 Two Divisions Case

In this section, we focus on the case where the firm owns only two divisions. For the firm, the optimization problem is then to decide on profit target t_1 and t_2 for the two divisions, respectively. We first have the following proposition.

Proposition 4. *Suppose that the firm with profit target T owns two divisions. Under the following reasonable assumptions:*

$$t_1 \le (b_2 - 1)t_2 + T \text{ and } t_2 \le (b_1 - 1)t_1 + T, \tag{15.30}$$

Fig. 15.1 The four possible scenarios where the firm achieves its profit target



the firm's profit probability is given by:

$$P_T(t_1,t_2) = P\{\varepsilon_1 \ge L_1 G_1^{-1}(K(t_1,t_2) - b_2 t_2), \varepsilon_2 \ge L_2 G_2^{-1}(K(t_1,t_2) - b_1 t_1),\ G_1 L_1^{-1}\varepsilon_1 + G_2 L_2^{-1}\varepsilon_2 \ge K(t_1,t_2)\}, \quad (15.31)$$

where

$$K(t_1,t_2) = (b_1-1)t_1 + (b_2-1)t_2 + T. \quad (15.32)$$

Proof. Based on (15.17), it can be seen that the firm's profit target T will be achieved if and only if

$$\sum_{i=1}^{2} G_i(H_i t_i - L_i^{-1}\varepsilon_i)^+ \le (t_1 + t_2 - T). \quad (15.33)$$

Of course, we should have the constraint $t_1 + t_2 \ge T$ to guarantee that T is achievable at all. Depending on the possible realizations of ε_1 and ε_2, we have the following four possible scenarios:

Scenario 1: If $\varepsilon_1 \ge L_1H_1t_1$ and $\varepsilon_2 \ge L_2H_2t_2$, (15.33) becomes $t_1 + t_2 \ge T$.
Scenario 2: If $\varepsilon_1 \le L_1H_1t_1$ and $\varepsilon_2 \le L_2H_2t_2$, (15.34) becomes $G_1L_1^{-1}\varepsilon_1 + G_2L_2^{-1}\varepsilon_2 \ge K(t_1,t_2)$.
Scenario 3: If $\varepsilon_1 \le L_1H_1t_1$ and $\varepsilon_2 \ge L_2H_2t_2$, (15.35) becomes $\varepsilon_1 \ge L_1G_1^{-1}[K(t_1,t_2) - b_2t_2]$.
Scenario 4: If $\varepsilon_1 \ge L_1H_1t_1$ and $\varepsilon_2 \le L_2H_2t_2$, (15.36) becomes $\varepsilon_2 \ge L_2G_2^{-1}[K(t_1,t_2) - b_1t_1]$.

The four scenarios are shown graphically in Fig. 15.1.

When plotting the graph, assumptions as in (15.30) are employed to guarantee that $L_1G_1^{-1}(K(t_1,t_2) - b_2t_2) > 0$ and $L_2G_2^{-1}(K(t_1,t_2) - b_1t_1) > 0$.

These assumptions are reasonable because a division's profit target is generally no more than the firm's profit target. Moreover, it can be verified that:

$$L_1 H_1 t_1 - L_1 G_1^{-1}(K(t_1,t_2) - b_2 t_2) = L_1 G_1^{-1}(t_1 + t_2 - T)$$
$$L_2 H_2 t_2 - L_2 G_2^{-1}(K(t_1,t_2) - b_1 t_1) = L_2 G_2^{-1}(t_1 + t_2 - T). \quad (15.34)$$

Hence, the Area 2 in Fig. 15.1 directly depends on the difference between the firm's profit target and the sum of the divisions' profit targets. Finally, because the firm's profit probability $P_T(t_1,t_2)$ is the sum of the four areas in Fig. 15.1, (15.31) is true. This concludes the proof. □

Proposition 4 greatly simplifies the calculation of $P_T(t_1,t_2)$ for the case of two divisions. To obtain more concrete results, now we assume two identical divisions. This means that the two divisions have the same production cost c, the same price elasticity b, and ε_1 and ε_2 are independent and identically distributed. Given these two identical divisions, it is practical to assume that the two divisions should be assigned an identical profit target, denoted by t. The optimization problem for the firm is then to select a single profit target t for both divisions so that the firm's profit probability is maximized. We have the following theorem.

Theorem 3. *Suppose that the firm with profit target T is in control of two identical divisions. Then its profit probability is given by:*

$$P_T(t) = P\{\varepsilon_1 + \varepsilon_2 \geq K_1(t), \varepsilon_1 \geq K_2(t), \varepsilon_2 \geq K_2(t)\}, \quad (15.35)$$

where

$$K_1(t) = LG^{-1}[(2b-2)t + T] \quad (15.36)$$
$$K_2(t) = LG^{-1}[(b-2)t + T]. \quad (15.37)$$

Furthermore, if $b \geq 2$, the optimal profit target for each division t^* is just half of the firm's profit target T, i.e., $t^* = T/2$.

Proof. Because this is the special case of two identical divisions, we can (15.35) readily from (15.31). In addition, the requirements in (15.30) become:

$$(2-b)t \leq T. \quad (15.38)$$

If $b \geq 2$, (15.38) is true. If $1 < b < 2$, (15.38) is true as well because $t \geq T/2$. Therefore, for the case of two for the case of two identical divisions, the assumption in (15.38) is satisfied automatically. To prove the second half of the theorem, we differentiate $P_T(t)$ with respect to t:

$$\begin{aligned} P'_T(T) = &-(b-2) f_{\varepsilon_1|\varepsilon_2 \geq k_2(t), \varepsilon_1+\varepsilon_2 \geq K_1(t)}(K_2(t)) P\{\varepsilon_2 \geq K_2(t), \varepsilon_1 + \varepsilon_2 \geq K_1(t)\} \\ &-(b-2) f_{\varepsilon_2|\varepsilon_1 \geq K_2(t), \varepsilon_1+\varepsilon_2 \geq K_1(t)}(K_2(t)) P\{\varepsilon_1 \geq K_2(t), \varepsilon_1 + \varepsilon_2 \geq K_1(t)\} \\ &-2(b-1) f_{\varepsilon_1+\varepsilon_2|\varepsilon_1 \geq K_2(t), \varepsilon_2 \geq K_2(t)}(K_2(t)) P\{\varepsilon_1 \geq K_2(t), \varepsilon_2 \geq K_2(t)\}. \end{aligned} \quad (15.39)$$

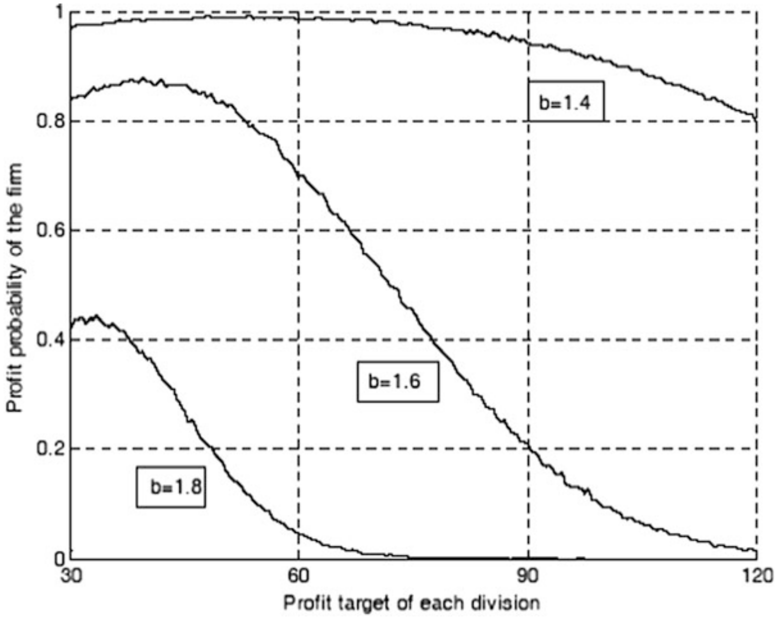


Fig. 15.2 The firm's profit probability as a function of each division's profit target when $b = 1.4, 1.6$, and 1.8

If the price elasticity $b \geq 2$, $P_T'(T) < 0$. Therefore, the firm's profit probability decreases in t and the optimal profit target t^* should be set at its minimum, i.e., $t^* = T/2$. This concludes the proof. □

It can be seen when the price elasticity of the product is reasonably large ($b \geq 2$), the optimal profit target for each division is always half of the firm's profit target. This result holds independent of all the other parameters including the underlying market demand distribution and production cost. It is also worth noting that for many individual products, it is common to have substitutes from competitors in the market (Gwartney 1976). As a result, $b \geq 2$ should hold for many business scenarios.

In the case of a relatively small price elasticity, in this case $b < 2$, it is difficult to obtain an analytic expression for the optimal profit target in general. In this case, the use of numerical simulation may be necessary. We present two examples here. In the first example, we conduct a computer simulation when the underlying market demand is normally distributed. In the second example, we are able to derive the closed-form expression of the optimal profit target for uniform distribution.

Example 2. Suppose that ε_i follows a normal distribution with mean 200 and standard deviation 50. Other parameters are: $c = 2$, $T = 60$, and $b = 1.4$, 1.6, and 1.8. Figure 15.2 shows the firm's profit probability as a function of the profit target of each division when $b = 1.4$, 1.6, and 1.8. The computer simulation is implemented using Matlab.

It can be seen from Fig. 15.2 that the firm's maximal profit probability ($P_T^*(t)$) is quite sensitive to the price elasticity b. As b changes from 1.4 to 1.8 (a change of 29%), $P_T^*(t)$ changes from almost 1.00 to approximately 0.45 (a change of 55%). Finally, the optimal target for each division (t^*) decreases with respect to b. As b approaches 2, the optimal target t^* approaches $T/2 = 30$.

Example 3. Suppose that ε_i follows a uniform distribution with lower bound 0 and upper bound β. To calculate the probability based on (15.31), we need to check the lower and upper bounds of $K_1(t)$ and $K_2(t)$ defined in (15.32) and (15.33), respectively. This is important because the random variable ε_i now is defined on a limited interval.

Because we have $T \leq 2t$, we have $K_2(t) = LG^{-1}[(b-2)t+T] \leq G^{-1}Lbt$. Further, based on Proposition 3 (see Sect. 15.2), we have $0 < K_2(t) < LG^{-1}b^*(g-c)L^{-1}\beta = \beta$. Similarly, we can have $0 < K_1(t) < 2\beta$.

Based on (15.35), we can have:

$$\begin{aligned} P_T(t) &= P\{\varepsilon_1 \geq K_2(t), \varepsilon_2 \geq K_2(t)\} - P\{\varepsilon_1 \geq K_2(t), \varepsilon_2 \geq K_2(t), \varepsilon_1 + \varepsilon_2 \leq K_1(t)\} \\ &= \frac{(\beta - K_2(t))^2}{\beta^2} - \frac{L^2}{2G^2}\frac{(2t-T)^2}{\beta^2}. \end{aligned} \tag{15.40}$$

The first- and second-order conditions are given by:

$$P_T'(t) = \frac{2L^2}{\beta^2 G^2}[(2-b)(L^{-1}G\beta + (2-b)t - T) - (2t-T)]. \tag{15.41}$$

$$P_T''(t) = \frac{2L^2}{\beta^2 G^2}\left[(2-b)^2 - 2\right] \tag{15.42}$$

When $1 < b < 2$, we have $(2-b)^2 - 2 < 0$ and, hence, $P_T''(t) < 0$. Therefore, the profit probability $P_T(t)$ is concave in the divisions' profit target t. The solution to $P_T'(t) = 0$ is given by:

$$t^o = \frac{(b-1)T + L^{-1}G(2-b)\beta}{2-(2-b)^2}. \tag{15.43}$$

Therefore, the optimal profit target for each division is given by $t^* = \max(t^o, T/2)$. Similar to Example 2, as the price elasticity approaches 2, t^o and, hence, the optimal target t^* approaches $T/2$.

15.5 Conclusions and Future Research

Existing studies have shown that the objective of maximizing profit probability leads to vastly different managerial insights than those based on the objective of maximizing expected profit. For example, Shi and Chen (2007) demonstrate that the simple wholesale price contract, when properly designed, can coordinate a supply chain with a single supplier and a single retailer, both of which adopt the objective of maximizing profit probability.

However, there has been little research on profit target setting under stochastic customer demand. We attempt to fill this gap with this chapter. To be more specific, we present analytic models on profit target setting under a common business scenario where a firm owns n divisions. While each division always has a profit target to achieve and, hence, adopts the objective of maximizing profit probability, the firm may adopt the objective of maximizing expected profit or the objective of maximizing profit probability. Given its assigned target, each division acts as a price-setting newsvendor and decides on divisional production level and selling price simultaneously.

We start by deriving the optimal behavior of a single division given its assigned profit target. We obtain the close-form expressions of the optimal production level, the optimal selling price, and the maximal profit probability. We show that for a division with a relatively higher (lower) production cost, a greater price elasticity always leads to a smaller (larger) profit probability.

We then study the problem of profit target setting from the firm's perspective. We first study the firm with the objective of maximizing expected profit, where we obtain closed-form expressions of the optimal profit target for each division. We then study the firm with the objective of maximizing profit probability. After deriving the firm's profit probability as a function of the divisions' profit targets, we proceed to study two special cases to gain more managerial insights.

In the first case of fair target setting, we derive the optimal profit targets when each division has a demand distribution with the property of increasing failure rate. We show that for a division with a relatively high (low) production cost, a greater price elasticity always leads to a (lower) higher profit target. In the second case where the firm has two divisions, we first derive the firm's profit probability as a function of the profit targets of the two divisions. If those divisions are identical, we show that each division's optimal target is just half of the firm's profit target when the price elasticity is two or more. This result is true regardless of the other parameters such as production cost and the demand distribution.

Our results in this chapter are helpful for practitioners who engage in the profit target setting and target attainment. However, much more work can be done in the area of profit target setting, which is both interesting and challenging. The first natural extension is to confirm the robustness of the results in this chapter by considering additive demand models. It is well documented that additive and multiple demand models may lead to different managerial insights (see, e.g., Choi 1991). The second natural extension is to see how risk seeking of firms impacts profit target setting for multiple divisions. Risk seeking can be modeled, for example, using the utility function approach where variance positively contributes to utility. Third, we can try to answer the following question: how does a firm set targets on multiple and potentially conflicting performance measures, such as profit, revenue, and market share? Fourth, we can study analytic target setting over multiple periods. For example, if a division of a firm has a yearly target to achieve, how can the division set quarterly targets and adjust them over time if necessary? Last but not least, it would be important to conduct empirical studies on profit target setting in business practice.

Acknowledgment This chapter is adapted from the following journal paper: The Setting of Profit Targets for Target Oriented Divisions. *European Journal of Operational Research,* 206, 86–92. We would like to thank Elsevier for granting permission. We would like to thank the handbook editor (Dr. Jason T. M. Choi) and two anonymous reviewers for their constructive suggestions.

References

Abbas, A. E., & Matheson, J. E. (2005). Normative target-based decision making. *Managerial and Decision Economics, 26*, 373–385.

Abbas, A. E., Matheson, J. E., & Bordley, R. F. (2009). Effective utility functions induced by organizational target-based incentives. *Managerial and Decision Economics, 30*, 235–251.

Aghion, P., & Tirole, J. (1997). Formal and real authority in organizations. *Journal of Political Economy, 105*, 1–29.

Agrawal, V., & Seshadri, S. (2000). Impact of uncertainty and risk-aversion on price and order quantity in the newsvendor problem. *Manufacturing and Service Operations Management, 2*, 410–423.

Barlow, R. E., & Proschan, F. (1965) *Mathematical theory of reliability*. New York: Wiley.

Bordley, R. F., & Kirkwood, C. W. (2004). Multiattribute preference analysis with performance targets. *Operations Research, 52*, 823–835.

Bouwens, J., & Van Lent, L. (2007). Assessing the performance of business unit managers. *Journal of Accounting Research, 45*(4), 667–697.

Brown, A., & Tang, C. S. (2006). The impact of alternative performance measures on single-period inventory policy. *Journal of Industrial and Management Optimization, 2*, 297–318.

Burgstahler, D., & Dichev, I. (1997). Earnings management to avoid earnings decreases and losses. *Journal of Accounting and Economics, 24*(1), 99–126.

Bushman, R. M., Indjejikian, R. J., & Smith, A. (1995). Aggregate performance measures in business unit manager compensation: the role of intrafirm interdependencies. *Journal of Accounting Research, 33*, 101–128.

Choi, S. C. (1991). Price competition in a channel structure with a common retailer. *Marketing Science, 10*, 271–296.

Dechow, P. M., & Skinner, D. J. (2000). Earnings management: reconciling the views of accounting academics, practitioners, and regulators. *Accounting Horizons, 14*(2), 235–250.

Degeorge, F., Patel, J., & Zeckhauser, R. (1999). Earnings management to exceed thresholds. *Journal of Business, 72*(1), 1–33.

Gan, X., Sethi, S. P., & Yan, H. (2005). Channel coordination with a risk-neutral supplier and a downside-risk-averse retailer. *Production and Operations Management, 14*(1), 80–89.

Giles W. (2008) Swiss Life to miss profit target, halts share buyback. Available: http://www.bloomberg.com/apps/news?pid=20601203&sid=a.GQDxalz0r8&refer=insurance. Accessed 12 Oct 2009.

Gwartney, J. (1976) *Economics: Private and public choice*. New York: Academic.

Healy, P. M., & Wahlen, J. M. (1999). A review of the earnings management literature and its implications for standard setting. *Accounting Horizons, 13*(4), 365–383.

Jensen, M. C. (2003). Paying people to lie: the truth about the budgeting process. *European Financial Management, 9*(3), 379–406.

Kabak, I. W., & Schiff, A. I. (1978). Inventory models and management objectives. *Sloan Management Review, 19*, 53–59.

Lariviere, M. A. (2006). A note on probability distributions with increasing generalized failure rates. *Operations Research, 54*, 602–604.

Lau, H. (1980). The newsboy problem under alternative optimization objectives. *Journal of the Operational Research Society, 31*, 393–403.

Lau, A., & Lau, H. (1988). Maximizing the probability of achieving a target profit level in a two-product newsboy problem. *Decision Sciences, 19*, 392–408.

Li, J., Lau, H., & Lau, A. (1990). Some analytic results for a two-product newsboy problem. *Decision Sciences, 21*, 710–725.

Li, J., Lau, H., & Lau, A. (1991). A two-product newsboy problem with satisficing objective and independent exponential demands. *IIE Transactions, 23*, 29–39.

Locke, E. A. (2001) Motivation by goal setting. In R. T. Golembiewski (Ed.), *Handbook of Organizational Behavior*. New York: Dekker.

Locke, E. A., & Latham, G. P. (1990) *A theory of goal setting and task performance*. Englewood Cliffs: Prentice Hall.

Locke, E. A., & Latham, G. P. (2002). Building a practically useful theory of goal setting and task motivation. *American Psychologist, 57*(9), 705–717.

Lovell, C. A. K., & Pastor, J. T. (1997). Target setting: an application to bank branch network. *European Journal of Operational Research, 98*, 290–299.

Merchant, K. A., & Manzoni, J. F. (1989). The achievability of budget targets in profit centers: a field study. *The Accounting Review, 3*, 539–558.

Monahan, G. E., Petruzzi, N. C., & Zhao, W. (2004). The dynamic pricing problem from a newsvendor's perspective. *Manufacturing and Service Operations Management, 6*, 73–91.

Ozler, A., Tan, B., & Karaesmen, F. (2009). Multi-product newsvendor problem with value-at-risk considerations. *International Journal of Production Economics, 117*, 244–255.

Parlar, M., & Weng, Z. K. (2003). Balancing desirable but conflicting objectives in the newsvendor problem. *IIE Transactions, 35*, 131–142.

Roberts, J. (2004) *The modern firm: organizational design for performance and growth*. Oxford: Oxford University.

Shi, C. V., & Chen, B. (2007). Pareto-optimal contracts for a supply chain with satisficing objectives. *Journal of the Operational Research Society, 58*, 751–760.

Yang, J. B., Wong, B. Y. H., Xu, D. L., & Stewart, T. J. (2009). Integrating DEA-oriented performance assessment and target setting using interactive MOLP methods. *European Journal of Operational Research, 195*, 205–222.

Chapter 16
A Portfolio Approach to Multi-product Newsvendor Problem with Budget Constraint*

Bin Zhang and Zhongsheng Hua

Abstract This chapter investigates a portfolio approach to multi-product newsvendor problem with budget constraint, in which the procurement strategy for each newsvendor product is designed as portfolio contract. A portfolio contract consists of a fixed-price contract and an option contract. We model the problem as an expected profit-maximization model, and propose an efficient solution procedure after investigating the structural properties of the model. We conduct numerical studies to show the efficiency of the proposed solution procedure, and to compare three models with different procurement contracts, i.e., fixed-price contract, option contract, and portfolio contract. Numerical results are shown to demonstrate the advantage of the portfolio model, and sensitivity analysis is provided for obtaining some managerial insights.

Keywords Newsvendor • Option contract • Portfolio • Budget constraint • Multiple products

16.1 Introduction

Multi-product constrained newsvendor problem is a classical inventory management problem, which was firstly introduced by Hadley and Whitin (1963). After Hadley and Whitin's seminal work, many researchers have investigated different models and

---

*The previous version of this work has been published in Computers & Industrial Engineering, 2010, 58: 759-765.

B. Zhang (✉)
Lingnan College, Sun Yat-Sen University, Guangzhou, 510275, PR China
e-mail: bzhang3@mail.ustc.edu.cn

Z. Hua
School of Management, University of Science and Technology of China,
Hefei, Anhui 230026, PR China
e-mail: zshua@ustc.edu.cn

T.-M. Choi (ed.), *Handbook of Newsvendor Problems*, International Series in Operations Research & Management Science 176, DOI 10.1007/978-1-4614-3600-3_16,

solution methods for multi-product newsvendor problems. Khouja (1999) presented a good literature review on the research. Due to the difficulty of solving large-scale problems, most recent works have focused on developing efficient solution methods, e.g. Lau and Lau (1996), Erlebacher (2000), Vairaktarakis (2000), Moon and Silver (2000), and Abdel-Malek et al. (2004). To address nonnegativity constraints of the order quantities, Abdel-Malek and Montanari (2005) proposed a modified Lagrangian-based method by analyzing the solution space. Zhang et al. (2009) provided an exact solution method for the problem with any continuous demand distribution. Zhang and Hua (2008) proposed a unified algorithm for solving a class of convex separable nonlinear knapsack problems, which include the singly constrained multi-product newsvendor problem with box constraints. Zhang and Du (2010) studied the multi-product newsvendor problem with limited capacity and outsourcing. Zhang (2012) analyzed structural properties of the multi-product newsvendor problem with multiple constraints, and proposed a multi-tier binary solution method for solving the exact solution. Zhang (2011) investigated a multi-product newsvendor problem with limited capacity in the presence of mixed deterministic and stochastic demands.

In the classical newsvendor problem, the product is procured from the supplier with fixed-price contract. Under this procurement strategy, the retailer will undertake the salvage loss resulting from lower realized demand. To avoid this risk, the retailer always does not order enough inventories to maximize the supply chain's total profit under the fixed-price contract (Cachon 2003). In order to maximize the supply chain's total profit, and share the risk raised from demand uncertainty with supply chain partners, some different contract types are used for encouraging the retailer to increase the order in supply chain management practice, such as buy back contracts, revenue sharing contracts, quantity flexibility contracts, sales rebate contracts, and quantity discount contracts (Cachon 2003). These contracts are labeled as "flexibility contract", in which a fixed amount of supply is determined when the contract is signed, but the amount to be delivered and paid for can differ from the quantity determined upon signing the contract. In comparison with fixed-price contracts, these flexibility contracts not only coordinate the supply chain, but also have sufficient flexibility (by adjusting parameters) to allow for any division of the supply chain's profit between suppliers and retailers. For more details of these flexibility contracts, please refer to Cachon (2003).

Option contract is one type of flexibility contracts (Martínez de Albéniz and Simchi-Levi 2005), which is defined as an agreement between the retailer and the supplier, in which the retailer pre-pays a reservation cost up-front for a commitment from the supplier to reserve certain order quantity. If the retailer does not execute the option, the initial payment is lost. With option contract, the retailer can purchase any amount of supply up to the option reservation level by paying an execution cost for each unit purchased. In other words, option contract provides the retailer with flexibility to adjust order quantity depending on the realized demand, and, hence, the inventory risk can be lowered for the retailer by utilizing the flexibility of option contract.

There are mainly two branches for the research on option contracts in supply chain management literature: One perspective is supply chain coordination with option contracts, e.g., Barnes-Schuster et al. (2002), Wu et al. (2002), Kleindorfer and Wu (2003), Wu and Kleindorfer (2005), Wang and Liu (2007), Gomez_Padilla and Mishina (2009), and Zhao et al. (2010). The other is on a single firm's optimal procurement decisions given particular contractual terms, e.g., Cohen and Agrawal (1999), Marquez and Blanchar (2004), Wang and Tsao (2006), and Boeckem and Schiller (2008), etc. In the research, some combinations of different contracts, such as fixed-price contract, and option contract, have been investigated.

In addition, some research on option contracts also took into account spot market since it is another source of supply for commodity products, e.g., Martínez de Albéniz and Simchi-Levi (2005), Aggarwal and Ganeshan (2007), and Fu et al. (2010). Spot market is a supply market in which products are sold for cash and delivered immediately. Contracts bought and sold on spot market are immediately effective. For some products, spot market can be used by the firm to purchase at any time; however, the product price on spot market is random. Over the last years, the emergence of the business-to-business trading exchange has transformed the procurement strategies, which provides spot market where buyers and sellers can trade products any time at online markets (Aggarwal and Ganeshan 2007). As Carbone (2001) reported, 50% of Hewlett-Packard's procurement cost was spent on fixed-price contract, 35% in option contracts, and the remaining was left to the spot market.

Up to now, all the existing works on the combination strategies of different contracts focused on single product setting. We have not found any research on multi-product demand management with the combination of fixed-price contract and option contract. In this chapter, we introduce a portfolio approach for managing multi-product newsvendor problem with budget constraint, in which each product can be procured with a portfolio contract consisting of a fixed-price contract and an option contract. The dual contracts for each product in the problem make the optimal ordering decisions more challenging in multi-product setting. On one hand, the use of option contract for lowering the overage cost should be properly balanced against the additional cost of using the option contract since unit reservation plus execution cost of option contract is typically higher than unit cost of a fixed-price contract. On the other hand, the total budget should be well allocated to different products for signing the fixed-price contracts and option contracts.

The overall objective of the newsvendor is to decide the optimal quantities of portfolio contracts for maximizing the total expected profit. We establish the structural properties for the optimal decisions of the proposed profit-maximization model, and develop an efficient solution procedure for the studied problem. Numerical results are shown to demonstrate the advantage of the portfolio model, and sensitivity analysis is provided for obtaining some managerial insights.

The rest of the chapter is organized as follows: Section 16.2 describes the problem formulation. In Sect. 16.3, the properties of the optimal solution are established, and an exact solution procedure is developed. Section 16.4 is dedicated to numerical studies for demonstrating the advantage of the portfolio contract model, and obtaining some managerial insights from sensitivity analysis. Section 16.5 briefly concludes the chapter and provides some future research directions.

16.2 Problem Formulation

We consider the following multi-product newsvendor problem. A retailer sells n different products with stochastic demands over a single period, and each product can be acquired from the suppliers by signing a portfolio contract, which includes a fixed-price contract and an option contract. In the fixed-price contract, the retailer pay unit fixed cost for procuring each product; in the option contract, the retailer pays unit reservation cost up-front for a commitment from the supplier, then the retailer can pays unit execution cost for procuring each product under the commitment level. If retailer does not exercise the option, the initial payment is lost. The retailer has limited budget for signing the portfolio contracts. In the following, we use i to be the index of product $1,\dots,n$.

The cost parameters used in this chapter are summarized in the following:

$p_i =$ Unit selling price for product i;
$s_i =$ Unit salvage value for product i;
$c_i =$ Unit procurement cost of fixed-price contract for product i;
$v_i =$ Unit reservation cost of option contract for product i;
$w_i =$ Unit execution cost of option contract for product i;
$B =$ Total budget available for signing the portfolio contracts.

To avoid the trivial case, we assume that $p_i > c_i > s_i$ for $i = 1,\dots,n$. Typically, the total cost of the option contract (reservation plus execution cost) is assumed to be larger than the cost of fixed-price contract, i.e., $v_i + w_i > c_i$; otherwise, the fixed-price contract is dominated by the option contract, and, hence, the fixed-price contract will never be engaged in the problem. We also assume that the reservation cost of option contract is smaller than the pure procurement cost of the fixed-price contract, i.e., $v_i < c_i - s_i$; otherwise, the option contract will be dominated by the fixed-price contract because the fixed-price contract always has the lower costs whether the product can be sold or not. From these two assumptions, i.e., $v_i + w_i > c_i$ and $v_i < c_i - s_i$, we have $s_i < w_i$, which implies that the retailer will not have an opportunity to make profit by executing an option contract in order to obtain the product salvage value.

The retailer makes quantity decisions of the portfolio contracts to fulfill n independent and stochastic demands. Let D_i denote the random demand for product $i = 1,\dots,n$, which has continuous probability density function $f_i(\cdot)$, cumulative distribution function $F_i(\cdot)$, and reverse distribution function $F_i^{-1}(\cdot)$. It is not uncommon to assume that all demands are nonnegative, thus, we can assume that $F_i(x) = 0$ for all $x < 0$, and $F_i(0) \geq 0$. This assumption does not rule out normal distribution as well as many other distributions with negative support values, since the distributions with negative support values should be approximated as nonnegative demand distributions in practice (Zhang and Du 2010).

The retailer's decisions are made in two stages. At the first stage, the retailer receives demand forecasts for all products, and determines a fixed-price contract quantity x_i, and an option contract quantity y_i to be signed. At the second stage, all demands are realized and the retailer exercises the quantity $\min((D_i - x_i)^+, y_i)$ of product i from the option contract to satisfy the demands for maximizing the revenue, where $(\cdot)^+ = \max\{\cdot, 0\}$.

We are ready to present profit-maximization model (denoted as problem P):

$$\text{Max}\pi(x,y) = \sum_{i=1}^{n} \pi_i = \sum_{i=1}^{n} E_i \begin{bmatrix} p_i \min(D_i, x_i + y_i) + s_i(x_i - D_i)^+ \\ -c_i x_i - v_i y_i - w_i \min((D_i - x_i)^+, y_i) \end{bmatrix}, \tag{16.1}$$

Subject to

$$\sum_{i=1}^{n} (c_i x_i + v_i y_i) \le B, \tag{16.2}$$

$$x_i \ge 0,\ y_i \ge 0,\ i = 1, \ldots, n. \tag{16.3}$$

For each product $i = 1, \ldots, n$, the first term $p_i \min(D_i, x_i + y_i)$ in (16.1) is the selling revenue, the second term $s_i(x_i - D_i)^+$ is the salvage value, the third term $c_i x_i$ is the acquisition cost with the fixed-price contract, the fourth term $v_i y_i$ is the option reservation cost, and the last term $w_i \min((D_i - x_i)^+, y_i)$ is the option execution cost. Equation (16.2) specifies the budget constraint on the quantities of the portfolio contracts. Note that the execution costs are excluded from the budget constraint because they are not needed to pay when signing the contracts at the first stage. Equation (16.3) gives the nonnegative constraints on the order quantities.

By using the formula $\min(D_i, x_i + y_i) = x_i + y_i - (x_i + y_i - D_i)^+$ and integration by parts formula $\int_0^x (x - z)\mathrm{d}F(z) = \int_0^x F(z)\mathrm{d}z$, the expected profit of problem P can be rewritten as:

$$\begin{aligned}
\pi(x,y) &= \sum_{i=1}^{n} E_i \begin{bmatrix} p_i(x_i + y_i - (x_i + y_i - D_i)^+) + s_i(x_i - D_i)^+ \\ -c_i x_i - v_i y_i - w_i(y_i - (x_i + y_i - D_i)^+ + (x_i - D_i)^+) \end{bmatrix} \\
&= \sum_{i=1}^{n} \begin{bmatrix} (p_i - c_i)x_i + (p_i - w_i - v_i)y_i \\ -(p_i - w_i)\int_0^{x_i + y_i} (x_i + y_i - z_i)\mathrm{d}F_i(z_i) - (w_i - s_i)\int_0^{x_i} (x_i - z_i)\mathrm{d}F_i(z_i) \end{bmatrix} \\
&= \sum_{i=1}^{n} \begin{bmatrix} (p_i - c_i)x_i + (p_i - w_i - v_i)y_i \\ -(p_i - w_i)\int_0^{x_i + y_i} F_i(z_i)\mathrm{d}z_i - (w_i - s_i)\int_0^{x_i} F_i(z_i)\mathrm{d}z_i \end{bmatrix}.
\end{aligned} \tag{16.4}$$

16.3 Properties and Solution Procedure

In this section, we first establish some structural properties for the optimal decisions, and then we develop an efficient solution method for the studied problem.

16.3.1 Properties of the Optimal Solution

Beginning with the objective function, we have the following proposition:

Proposition 1. *The expected profit function π is jointly concave in x_i and $y_i, i = 1,\ldots,n$.*

Proof. Since

$$\begin{cases} \partial\pi/\partial x_i = (p_i - c_i) - (p_i - w_i)F_i(x_i + y_i) - (w_i - s_i)F_i(x_i) \\ \partial\pi/\partial y_i = (p_i - w_i - v_i) - (p_i - w_i)F_i(x_i + y_i) \end{cases}, \quad i = 1,\ldots,n,$$

we have

$$\begin{cases} \partial^2\pi/\partial x_i^2 = -(p_i - w_i)f_i(x_i + y_i) - (w_i - s_i)f_i(x_i) \le 0 \\ \partial^2\pi/\partial y_i^2 = \partial^2\pi/\partial x_i\partial y_i = -(p_i - w_i)f_i(x_i + y_i) \le 0 \end{cases}, \quad i = 1,\ldots,n,$$

and $\partial^2\pi/\partial x_i\partial x_j = \partial^2\pi/\partial y_i\partial y_j = \partial^2\pi/\partial x_i\partial y_j = 0$ for $i \ne j$, $i, j = 1,\ldots,n$. Thus, the Hessian matrix of the objective function is negative semi-definite. □

Since π is concave and the feasible domain of the problem is convex, the Karush–Kuhn–Tucker (KKT) conditions are necessary and sufficient for optimality. Let $\lambda \ge 0$, $\alpha_i \ge 0$, and $\beta_i \ge 0$, $i = 1,\ldots,n$, be the dual variables corresponding to the constraints in (16.2)–(16.3), respectively. Then (x_i, y_i), $i = 1,\ldots,n$, is optimal if and only if there exists nonnegative dual variables λ, α_i, and β_i, $i = 1,\ldots,n$, such that

$$(p_i - c_i) - (p_i - w_i)F_i(x_i + y_i) - (w_i - s_i)F_i(x_i) - \lambda c_i + \alpha_i = 0, \quad i = 1,\ldots,n, \tag{16.5}$$

$$(p_i - w_i - v_i) - (p_i - w_i)F_i(x_i + y_i) - \lambda v_i + \beta_i = 0,\ i = 1,\ldots,n, \tag{16.6}$$

$$\sum_{i=1}^{n} (\alpha_i x_i + \beta_i y_i) = 0, \tag{16.7}$$

$$\lambda\left(B - \sum_{i=1}^{n} (c_i x_i + v_i y_i)\right) = 0. \tag{16.8}$$

To solve these KKT conditions, we first investigate how to solve (16.5)–(16.7) with any given $\lambda \ge 0$, and then we illustrate how to decide the optimal value for λ. Denote by $(\tilde{x}_i^\lambda, \tilde{y}_i^\lambda, \tilde{\alpha}_i^\lambda, \tilde{\beta}_i^\lambda, \lambda)$, $i = 1,\ldots,n$, an solution of (16.5)–(16.7), then we have the following propositions:

Proposition 2. *For any given $\lambda \ge 0, (\tilde{x}_i^\lambda, \tilde{y}_i^\lambda), i = 1,\ldots,n$, satisfies*

$$\begin{cases} \tilde{x}_i^\lambda = F_i^{-1}\left(\min\left\{\left(\frac{w_i - (1+\lambda)(c_i - v_i)}{w_i - s_i}\right)^+, \left(\frac{p_i - (1+\lambda)c_i}{p_i - s_i}\right)^+\right\}\right) \\ \tilde{y}_i^\lambda = \left(F_i^{-1}\left(\left(\frac{p_i - w_i - (1+\lambda)v_i}{p_i - w_i}\right)^+\right) - F_i^{-1}\left(\left(\frac{w_i - (1+\lambda)(c_i - v_i)}{w_i - s_i}\right)^+\right)\right)^+ \end{cases}.$$

Proof. The proof of this proposition is presented in Appendix.

This proposition characterizes the optimal solution of (16.5)–(16.7) with any given $\lambda \geq 0$, and also indicates the optimal solution to the problem without budget constraint (denoted as problem P1). Denote by $(\tilde{x}_i, \tilde{y}_i)$, $i = 1, \ldots, n$, an optimal solution to problem P1, by simply setting $\lambda = 0$ in Proposition 2, then the optimal values of $(\tilde{x}_i, \tilde{y}_i)$, $i = 1, \ldots, n$ are given as follows:

$$\begin{cases} \tilde{x}_i = F_i^{-1}\left(\min\left\{\frac{w_i+v_i-c_i}{w_i-s_i}, \frac{p_i-c_i}{p_i-s_i}\right\}\right) \\ \tilde{y}_i = \left(F_i^{-1}\left(\frac{p_i-w_i-v_i}{p_i-w_i}\right) - F_i^{-1}\left(\frac{w_i+v_i-c_i}{w_i-s_i}\right)\right)^+ . \end{cases} \tag{16.9}$$

From the result of Proposition 2, we also know that the difference between the optimal solution of the constraint problem and that of the unconstraint problem increases too, when λ increases.

Before discussing the result in (16.9), we first investigate the optimal unconstrained orders under pure fixed-price contract (FC) and pure option contract (OC). If there is no budget constraint and only fixed-price contract is used, then we can solve the optimal unconstrained order $\tilde{x}_{i,\mathrm{FC}}$ from

$$\partial\pi/\partial x_i = (p_i - c_i) - (p_i - w_i)F_i(x_i + y_i) - (w_i - s_i)F_i(x_i) = 0,$$

by setting $y_i = 0$, and, hence, $\tilde{x}_{i,\mathrm{FC}} = F_i^{-1}\left(\frac{p_i-c_i}{p_i-s_i}\right)$, $i = 1, \ldots, n$. If there is no budget constraint and only pure option contract is used, then we can solve the optimal unconstrained order $\tilde{y}_{i,\mathrm{OC}}$ from

$$\partial\pi/\partial y_i = (p_i - w_i - v_i) - (p_i - w_i)F_i(x_i + y_i) = 0,$$

by setting $x_i = 0$, and, hence, $\tilde{y}_{i,\mathrm{OC}} = F_i^{-1}\left(\frac{p_i-w_i-v_i}{p_i-w_i}\right)$, $i = 1, \ldots, n$. Note that the option contract (v_i, w_i) can also be viewed as a fixed-price contract with unit purchase cost $c_i' = v_i + w_i$, and unit salvage value $s_i' = w_i$. Thus, $\tilde{y}_{i,\mathrm{OC}}$ and $\tilde{x}_{i,\mathrm{FC}}$ have the same form, i.e., $\tilde{y}_{i,\mathrm{OC}} = F_i^{-1}\left(\frac{p_i-w_i-v_i}{p_i-w_i}\right) = F_i^{-1}\left(\frac{p_i-c_i'}{p_i-s_i'}\right)$.

Let us discuss the relationship among the unconstrained solution $\tilde{x}_{i,\mathrm{FC}}$, $\tilde{y}_{i,\mathrm{OC}}$, and the optimal unconstrained order of portfolio contract, $(\tilde{x}_i, \tilde{y}_i)$ presented in (16.9). If $\frac{p_i-w_i-v_i}{p_i-w_i} > \frac{p_i-c_i}{p_i-s_i}$, then the mathematical transform gives $\frac{p_i-w_i-v_i}{p_i-w_i} > \frac{p_i-c_i}{p_i-s_i} > \frac{v_i+w_i-c_i}{w_i-s_i}$, and we have $\tilde{x}_i = F_i^{-1}\left(\frac{w_i+v_i-c_i}{w_i-s_i}\right)$ and $\tilde{y}_i = F_i^{-1}\left(\frac{p_i-w_i-v_i}{p_i-w_i}\right) - F_i^{-1}\left(\frac{w_i+v_i-c_i}{w_i-s_i}\right)$; thus, the total order quantity of portfolio contract is $\tilde{x}_i + \tilde{y}_i = F_i^{-1}\left(\frac{p_i-w_i-v_i}{p_i-w_i}\right)$; otherwise, we have $\frac{p_i-w_i-v_i}{p_i-w_i} \leq \frac{p_i-c_i}{p_i-s_i} \leq \frac{v_i+w_i-c_i}{w_i-s_i}$, and, hence, $\tilde{x}_i = F_i^{-1}\left(\frac{p_i-c_i}{p_i-s_i}\right)$ and $\tilde{y}_i = 0$; thus the total order quantity of portfolio contract is $\tilde{x}_i + \tilde{y}_i = F_i^{-1}\left(\frac{p_i-c_i}{p_i-s_i}\right)$. Therefore, the total optimal unconstrained order of portfolio contract can be expressed as

$$\tilde{x}_i + \tilde{y}_i = \max\left\{F_i^{-1}\left(\frac{p_i - c_i}{p_i - s_i}\right), F_i^{-1}\left(\frac{p_i - w_i - v_i}{p_i - w_i}\right)\right\} = \max\left\{\tilde{x}_{i,\mathrm{FC}}, \tilde{y}_{i,\mathrm{OC}}\right\}.$$

From the proof of Proposition 2, we have

$$\tilde{x}_i^{\lambda} + \tilde{y}_i^{\lambda} = \max\left\{ F_i^{-1}\left(\frac{p_i-(1+\lambda)c_i}{p_i-s_i}\right)^+, F_i^{-1}\left(\frac{p_i-w_i-(1+\lambda)v_i}{p_i-w_i}\right)^+ \right\}.$$

Since $\lambda \geq 0$, we know $\tilde{x}_i^{\lambda} + \tilde{y}_i^{\lambda} \leq \tilde{x}_i + \tilde{y}_i$, which means that the total optimal unconstrained order of portfolio contract is an upper bound for the total optimal order of portfolio contract in problem P. Thus, the maximum of the optimal unconstrained order under pure fixed-price contract and the optimal unconstrained order under pure option contract is an upper bound for the optimal total order of portfolio contract in problem P.

Denote by (x_i^*, y_i^*), $i = 1,\dots,n$, an optimal solution to problem P, and λ^* the corresponding optimal value of λ, we have the following proposition:

Proposition 3. *(a) If* $\sum_{i=1}^n (c_i\tilde{x}_i + v_i\tilde{y}_i) \leq B$, *then* $x_i^* = \tilde{x}_i$ *and* $y_i^* = \tilde{y}_i$, $i = 1,\dots,n$; *(b) If* $\sum_{i=1}^n (c_i\tilde{x}_i + v_i\tilde{y}_i) > B$, *then* $\sum_{i=1}^n (c_i x_i^* + v_i y_i^*) = B$.

Proof. (a) This property is obvious since the budget constraint is not active. It is also easily verified that $(x_i^*, y_i^*) = (\tilde{x}_i, \tilde{y}_i)$ with $\lambda^* = 0$ satisfy the condition in (16.8).

(b) If $\sum_{i=1}^n (c_i x_i^* + v_i y_i^*) < B$, according to $\sum_{i=1}^n (c_i\tilde{x}_i + v_i\tilde{y}_i) > B$, there must exist at least one $k \in \{1,\dots,n\}$ such that $c_k x_k^* + v_k y_k^* < c_k\tilde{x}_k + v_k\tilde{y}_k$. Since $\sum_{i=1}^n (c_i x_i^* + v_i y_i^*) < B$, the slackness condition $\lambda(B - \sum_{i=1}^n (c_i x_i + v_i y_i)) = 0$ in (16.8) implies $\lambda^* = 0$, and this further means $(x_i^*, y_i^*) = (\tilde{x}_i, \tilde{y}_i)$, $i = 1,\dots,n$, which violates $c_k x_k^* + v_k y_k^* < c_k\tilde{x}_k + v_k\tilde{y}_k$. Thus, we have $\sum_{i=1}^n (c_i x_i^* + v_i y_i^*) = B$. □

Property (a) indicates that the optimal solution to problem P is the same as that of problem P1 if budget constraint is inactive. Property (b) illustrates that the budget must be fully utilized at the optimal solution if the budget constraint is binding, i.e., $\sum_{i=1}^{n} (c_i\tilde{x}_i + v_i\tilde{y}_i) > B$.

16.3.2 Solution Procedure

Before developing the solution procedure, we first prove the following result:

Proposition 4. $c_i\tilde{x}_i^{\lambda} + v_i\tilde{y}_i^{\lambda}$ *is nonincreasing in* λ.

Proof. The proof of this proposition is presented in Appendix.

Fig. 16.1 Main steps of Algorithm 1

Step 0: Calculate $(\tilde{x}_i, \tilde{y}_i)$, $i = 1, \ldots, n$, from Eq (16.9).

Step 1: If $\sum_{i=1}^{n}(c_i\tilde{x}_i + v_i\tilde{y}_i) \le B$, then

let $(x_i^*, y_i^*) = (\tilde{x}_i, \tilde{y}_i)$, $i = 1, \ldots, n$, goto *Step* 6.

Step 2: Let $\lambda_L = 0$, $\lambda_U = \max_{i=1,\ldots,n}\left(\frac{p_i - c_i}{c_i}, \frac{p_i - w_i - v_i}{v_i}\right)$.

Step 3: Let $\lambda = (\lambda_L + \lambda_U)/2$;

Calculate $\tilde{x}_i^\lambda$ and $\tilde{y}_i^\lambda$, $i = 1, \ldots, n$, from Proposition 2.

Step 4: If $\sum_{i=1}^{n}(c_i\tilde{x}_i^\lambda + v_i\tilde{y}_i^\lambda) < B$, then let $\lambda_U = \lambda$, goto *Step* 3;

If $\sum_{i=1}^{n}(c_i\tilde{x}_i^\lambda + v_i\tilde{y}_i^\lambda) > B$, then let $\lambda_L = \lambda$, goto *Step* 3.

Step 5: Let $x_i^* = \tilde{x}_i^\lambda$ and $y_i^* = \tilde{y}_i^\lambda$, $i = 1, \ldots, n$.

Step 6: Output x_i^* and y_i^*, $i = 1, \ldots, n$, stop.

Proposition 4 provides a good property with which the optimal value of λ can be found without using any linear search method. We can develop an efficient way to decide the optimal value for λ when the budget constraint is binding.

If the budget constraint is binding, according to Proposition 3(b), we know that $\sum_{i=1}^{n}(c_i x_i^* + v_i y_i^*) = B$, which implies $\lambda^* \le \bar{\lambda} \equiv \max_{i=1,\ldots,n}\left(\frac{p_i - c_i}{c_i}, \frac{p_i - w_i - v_i}{v_i}\right)$. Otherwise, when $\lambda^* > \max\left\{\frac{p_i}{c_i} - 1, \frac{p_i - w_i}{v_i} - 1\right\}$, $i = 1, \ldots, n$, from the proof of Proposition 2, we know $\tilde{x}_i^{\lambda^*} = \tilde{y}_i^{\lambda^*} = 0$, $i = 1, \ldots, n$, which violates the necessary condition $\sum_{i=1}^{n}(c_i x_i^* + v_i y_i^*) = B$. Thus, according to the results in Propositions 2–4, we can determine λ^* by applying a binary search method over the interval $\lambda \in [0, \bar{\lambda}]$, and simultaneously solve the optimal solution to problem P.

Main steps of the solution procedure for solving the optimal solution to problem P are summarized in Algorithm 1 (as shown in Fig. 16.1).

In Algorithm 1, we first solve problem P1 (*Step* 0) to obtain $(\tilde{x}_i, \tilde{y}_i)$, $i = 1, \ldots, n$. Then we judge whether $(\tilde{x}_i, \tilde{y}_i)$, $i = 1, \ldots, n$, leads to a binding budget constraint or not (*Step* 1). If the budget constraint is inactive at $(\tilde{x}_i, \tilde{y}_i)$, $i = 1, \ldots, n$, then we let $(x_i^*, y_i^*) = (\tilde{x}_i, \tilde{y}_i)$, $i = 1, \ldots, n$. Otherwise, we apply the binary search procedure over interval $[\lambda_L, \lambda_U]$ to determine (x_i^*, y_i^*) $(i = 1, \ldots, n)$ (*Steps* 2–5). *Step* 6 outputs the optimal solution to problem P. Since we do not assume any specific property on demand distribution, our approach is applicable to any continuous demand distribution.

The computational complexity of Algorithm 1 is analyzed as follows. The complexity of Steps 0–2 is $O(n)$. The search of λ within the interval $[0, \bar{\lambda}]$ in Steps 3–5 needs $\log_2(\bar{\lambda}/\varepsilon)$ iterations, where ε is the error target for the binary search procedure. Take $\bar{\lambda} = 10^{10}$ and $\varepsilon = 10^{-6}$ as an example, the number of iterations for determining λ is $\log_2(\bar{\lambda}/\varepsilon) = 36.8414 \approx 37$. The computation procedure in each step of Steps 3–5 has complexity$O(n)$. So the computational complexity of Steps 3–5 is $O(\log_2(\bar{\lambda}/\varepsilon)n)$. The complexity of Step 6 is $O(n)$. Thus, Algorithm 1 has computational complexity $O(\log_2(\bar{\lambda}/\varepsilon)n)$, which is polynomial in the number of products.

16.4 Numerical Studies

In this section, numerical results are provided to show the efficiency of the proposed solution procedure, and to compare three models with different procurement contracts, i.e., fixed-price contract (FC), option contract (OC), and portfolio contract (PC). Sensitivity analysis is also provided for obtaining some managerial insights. The two pure contract models (i.e., fixed-price contract, pure option contract) are easily obtained from the portfolio contract model by setting $y_i = 0$ or $x_i = 0$, $i = 1, \ldots, n$, respectively. The portfolio contract model should dominate the two pure contract models since the optimal solutions to the two pure contract models are both feasible solutions to the portfolio contract model.

Before presenting numerical results, we first briefly illustrate how to solve pure fixed-price contract model and pure option contract model. The two pure contract models can be reformulated as minimizing $-\pi_{\mathrm{FC}}(x)$ and $-\pi_{\mathrm{OC}}(y)$, respectively:

$$(\mathrm{FC})\ \mathrm{Min} - \pi_{\mathrm{FC}}(x) = -\left[\sum_{i=1}^{n}(p_i - c_i)x_i - (p_i - s_i)\int_0^{x_i} F_i(z_i)\mathrm{d}z_i\right],$$

$$s.t.\ \sum_{i=1}^{n} c_i x_i \le B,\ x_i \ge 0,\ i = 1, \ldots, n.$$

$$(\mathrm{OC})\ \mathrm{Min} - \pi_{\mathrm{OC}}(y) = -\left[\sum_{i=1}^{n}(p_i - w_i - v_i)y_i - (p_i - w_i)\int_0^{y_i} F_i(z_i)\mathrm{d}z_i\right],$$

$$s.t.\ \sum_{i=1}^{n} v_i y_i \le B,\ y_i \ge 0,\ i = 1, \ldots, n.$$

Since the two pure contract models are the special cases of problem P where $y_i = 0$ or $x_i = 0$, $i = 1, \ldots, n$, according to Proposition 1, we know $\pi_{\mathrm{FC}}(x)$ is concave in x, and $\pi_{\mathrm{OC}}(y)$ is concave in y. Thus the objective functions of PC and OC models, i.e., $-\pi_{\mathrm{FC}}(x)$ and $-\pi_{\mathrm{OC}}(y)$, are both convex. It is obvious that the objective functions of PC and OC models are both separable, therefore the two pure contract models can be viewed as the class of convex separable nonlinear knapsack problems studied by Zhang and Hua (2008).

These knapsack problems have two important characteristics: positive marginal cost (PMC) and increasing marginal loss-cost ratio (IMLCR). PMC means that the budget occupancy increases in order quantity, which is guaranteed by the positive

Table 16.1 Parameters and solutions for the illustrative example

i	p_i	s_i	c_i	v_i	w_i	μ_i	σ_i	$x^*_{i,\mathrm{FC}}$	$y^*_{i,\mathrm{OC}}$	$x^*_{i,\mathrm{PC}}$	$y^*_{i,\mathrm{PC}}$
1	96	14	43	11	42	109	22	85.07	127.23	0.00	117.63
2	96	11	41	14	49	103	28	77.34	117.85	63.15	39.28
3	92	14	47	14	46	102	24	0.00	114.29	0.00	100.84
4	91	17	46	14	46	108	29	34.94	122.29	0.00	105.76
5	99	18	49	14	43	103	27	52.69	121.21	0.00	108.00
6	105	19	45	24	45	106	27	82.37	112.84	93.99	0.00
7	105	18	44	22	49	109	24	89.74	115.53	99.37	0.00
8	104	14	42	22	44	106	22	90.23	113.50	96.28	2.68
9	109	16	46	24	46	108	22	89.34	114.67	95.33	4.20
10	101	17	47	24	44	101	22	72.90	105.38	80.56	7.73
π^*								37,073.94	33,340.71	44,301.65	

linear constraint, i.e., $c_i > 0$, $i = 1,\ldots,n$, for FC model, and $v_i > 0$, $i = 1,\ldots,n$, for OC model; IMLCR requires that the ratio of marginal loss to marginal cost increases in order quantity, i.e. $\frac{\mathrm{d}[-\pi_{\mathrm{FC}}(x)]}{c_i \mathrm{d}x_i} = -\frac{p_i - c_i}{c_i} + \frac{p_i - s_i}{c_i} F_i(x_i)$ increases in x_i, $i = 1,\ldots,n$, for FC model, and $\frac{\mathrm{d}[-\pi_{\mathrm{OC}}(y)]}{v_i \mathrm{d}y_i} = -\frac{p_i - w_i - v_i}{v_i} + \frac{p_i - w_i}{v_i} F_i(y_i)$ increases in y_i, $i = 1,\ldots,n$, for OC model. Since FC and OC models satisfy PMC and IMLCR, the method with linear computation complexity developed by Zhang and Hua (2008) can be directly used to solve them.

It is worth noticing that the method proposed by Zhang and Hua (2008) is not available for solving the portfolio contract model proposed in this chapter, due to the fact that the objective function of problem P is nonseparable.

In the numerical experiment, the relative profit differences between FC, OC, and PC, i.e., $\Delta\pi^{\mathrm{PC}}_{\mathrm{FC}} = (\pi^*_{\mathrm{PC}} - \pi^*_{\mathrm{FC}})/\pi^*_{\mathrm{FC}} \times 100\%$, $\Delta\pi^{\mathrm{PC}}_{\mathrm{OC}} = (\pi^*_{\mathrm{PC}} - \pi^*_{\mathrm{OC}})/\pi^*_{\mathrm{OC}} \times 100\%$ are reported to show the benefit of portfolio contract model and to obtain some managerial insights through sensitivity analysis.

16.4.1 An Illustrative Example

In this example, demands of 10 products are all normally distributed, and there is a budget constraint $B = 30,000$. Table 16.1 shows the relevant information, where μ_i, σ_i, $i = 1,\ldots,n$, are parameters of the mean and standard deviation of the normal demand $x^*_{i,\mathrm{FC}}$ and $y^*_{i,\mathrm{OC}}$ are the optimal solutions of the fixed-price contract model and the option contract model, respectively; $x^*_{i,\mathrm{PC}}$ and $y^*_{i,\mathrm{PC}}$ are the optimal solutions of the portfolio contract model and π^* stands for the optimal expected profits of the three different models. In order to investigate the general case, we set the parameters such that $\frac{p_i - w_i - v_i}{p_i - w_i} > \frac{p_i - c_i}{p_i - s_i}$ for $i = 1,\cdots,5$ and $\frac{p_i - w_i - v_i}{p_i - w_i} < \frac{p_i - c_i}{p_i - s_i}$ for $i = 6,\cdots,10$. The results in Table 16.1 show that the portfolio contract model is better than the fixed-price contract and option contract models.

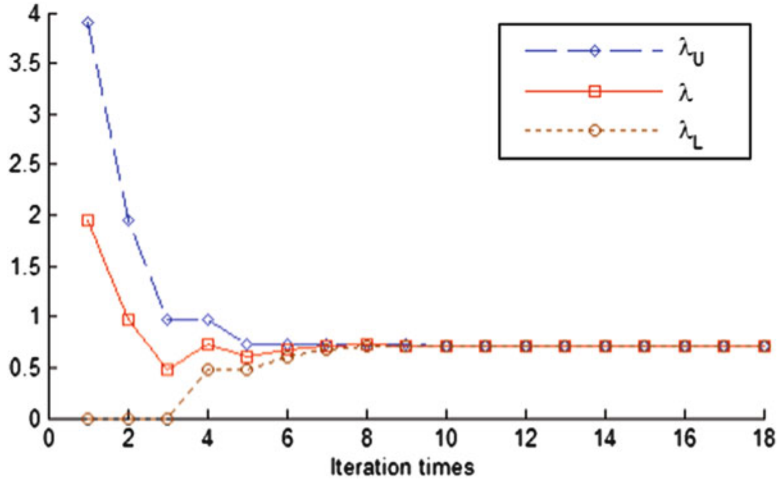


Fig. 16.2 The values of λ_L, λ_U, and λ in the iteration process

To show the efficiency of the proposed solution procedure, we plot the iterative solution process of this example in Fig. 16.2. In this figure, we report the values of λ_L, λ_U, and λ in the iteration process. From this figure, it can be observed that Algorithm 1 solves the optimal value $\lambda^* = 0.7059$ within only 18 iteration times.

16.4.2 Sensitivity Analysis

To investigate how the budget constraint affects the relative profit differences among three procurement strategies, we provide sensitivity analysis by changing B of the base example shown in Table 16.1 and keeping other parameters unchanged. The results of more cases with different B are presented in Tables 16.2–16.4. In Table 16.2, we report the optimal profits and the relative profit differences of different models, and the ratio of total budget used on option contracts, which is defined as $\Delta B^{\mathrm{O}}_{\mathrm{PC}} = \sum_{i=1}^{n} v_i y^*_{i,\mathrm{PC}} / \sum_{i=1}^{n} \left(c_i x^*_{i,\mathrm{PC}} + v_i y^*_{i,\mathrm{PC}}\right) \times 100\%$. Tables 16.3 and 16.4 give the optimal order quantity of fixed-price contract and the optimal order quantity of option contract in the portfolio contract for different B, respectively.

From Table 16.2, we have the following observations:

(1) The optimal profits of three procurement strategies are all nondecreasing in the available budget. This observation is obvious since a larger B will provide a larger feasible domain of the optimization problem.

Table 16.2 The profit comparisons, shadow prices, and ΔB_{PC}^{O} for different B

B	π_{FC}^{*}	π_{OC}^{*}	π_{PC}^{*}	$\Delta\pi_{FC}^{PC}(\%)$	$\Delta\pi_{OC}^{PC}(\%)$	λ_{FC}^{*}	λ_{OC}^{*}	λ_{PC}^{*}	$\Delta B_{PC}^{O}(\%)$
100	147.62	390.91	390.91	164.81	0.00	1.48	3.91	3.91	100
1,100	1,623.74	4,072.63	4,072.63	150.82	0.00	1.48	3.00	3.00	100
2,100	3,096.56	6,943.82	6,943.82	124.24	0.00	1.46	2.45	2.45	100
3,100	4,520.40	9,272.62	9,272.62	105.13	0.00	1.39	2.28	2.28	100
4,100	5,906.56	11,517.79	11,517.79	95.00	0.00	1.39	2.21	2.21	100
5,000	7,150.73	13,453.37	13,453.37	88.14	0.00	1.38	2.05	2.05	100
10,000	13,934.67	21,831.71	21,831.71	56.67	0.00	1.33	1.54	1.54	100
15,000	20,472.33	28,993.59	28,993.59	41.62	0.00	1.24	1.30	1.30	100
20,000	26,527.46	33,198.01	34,978.14	31.86	5.36	1.15	0.27	1.13	60.37
25,000	32,091.06	**33,340.71**	40,180.57	25.21	20.52	1.02	**0.00**	0.91	36.03
30,000	37,073.94	33,340.71	44,301.65	19.50	32.88	0.97	0.00	0.71	21.98
35,000	41,808.94	33,340.71	46,892.01	12.16	40.64	0.91	0.00	0.41	15.77
40,000	45,933.08	33,340.71	48,692.49	6.01	46.05	0.72	0.00	0.31	9.78
45,000	48,838.10	33,340.71	49,983.31	2.34	49.92	0.43	0.00	0.21	5.37
50,000	50,185.62	33,340.71	50,581.42	0.79	51.71	0.11	0.00	0.01	2.71
55,000	**50,276.22**	33,340.71	**50,582.34**	**0.61**	**51.71**	**0.00**	0.00	**0.00**	**2.63**
60,000	50,276.22	33,340.71	50,582.34	0.61	51.71	0.00	0.00	0.00	2.63

Table 16.3 The optimal order quantity of fixed-price contract in the portfolio contract

	$x_{i,PC}^{*}$									
B	1	2	3	4	5	6	7	8	9	10
100	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
1,100	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
2,100	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
3,100	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
4,100	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
5,000	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
10,000	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
15,000	0.00	0.00	0.00	0.00	0.00	**0.00**	**0.00**	**0.00**	0.00	0.00
20,000	0.00	0.00	0.00	0.00	0.00	40.74	73.00	68.60	**0.00**	**0.00**
25,000	0.00	**0.00**	0.00	0.00	0.00	81.87	90.72	86.80	83.27	17.89
30,000	0.00	63.15	0.00	**0.00**	0.00	93.99	99.37	96.28	95.33	80.56
35,000	**0.00**	87.44	**0.00**	54.62	0.00	104.94	108.70	105.92	106.66	96.26
40,000	34.84	92.75	68.96	76.45	**0.00**	108.33	111.61	108.39	109.27	99.24
45,000	82.66	97.97	80.78	88.40	49.67	111.94	114.72	111.02	112.04	102.36
50,000	100.31	108.10	95.71	105.99	89.37	119.60	121.34	116.57	117.85	108.75
55,000	100.95	108.58	96.31	106.75	90.37	119.98	121.67	116.84	118.13	109.05
60,000	100.95	108.58	96.31	106.75	90.37	119.98	121.67	116.84	118.13	109.05

(2) The optimal profits of three procurement strategies do not change when the available budget exceeds the maximal active budgets, which are $\sum_{i=1}^{n} c_i\tilde{x}_{i,\mathrm{FC}}$, $\sum_{i=1}^{n} v_i\tilde{y}_{i,\mathrm{OC}}$, and $\sum_{i=1}^{n} c_i\tilde{x}_{i,PC} + v_i\tilde{y}_{i,PC}$ in FC, OC, and PC models, respectively. The budget constraint is inactive when the budget is larger than the maximal

Table 16.4 The optimal order quantity of option contract in the portfolio contract

	$y^*_{i,\mathrm{PC}}$									
B	1	2	3	4	5	6	7	8	9	10
100	9.09	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
1,100	89.31	0.00	0.00	0.00	8.40	0.00	0.00	0.00	0.00	0.00
2,100	97.30	0.00	0.00	0.00	73.55	0.00	0.00	0.00	0.00	0.00
3,100	99.37	46.36	18.83	0.00	78.16	0.00	0.00	0.00	0.00	0.00
4,100	100.31	55.35	54.38	24.22	80.08	0.00	0.00	0.00	0.00	0.00
5,000	102.20	65.69	66.87	60.58	83.70	0.00	0.00	0.00	0.00	0.00
10,000	108.01	83.45	83.95	84.50	93.63	0.00	34.44	73.13	67.01	0.00
15,000	110.75	89.54	89.45	91.52	97.92	**68.17**	**77.83**	**83.87**	82.62	60.31
20,000	112.63	93.31	92.80	95.75	100.74	36.89	12.33	20.25	**88.47**	**73.03**
25,000	115.16	**98.07**	97.02	101.02	104.47	4.61	1.98	7.58	11.33	64.16
30,000	117.63	39.28	100.84	**105.76**	108.00	0.00	0.00	2.68	4.20	7.73
35,000	**121.38**	21.26	**106.33**	57.90	113.26	0.00	0.00	0.00	0.00	0.00
40,000	87.82	18.01	39.16	38.28	**115.02**	0.00	0.00	0.00	0.00	0.00
45,000	41.40	15.00	29.27	28.70	67.27	0.00	0.00	0.00	0.00	0.00
50,000	26.77	9.53	18.38	16.06	31.64	0.00	0.00	0.00	0.00	0.00
55,000	26.28	9.28	17.98	15.54	30.84	0.00	0.00	0.00	0.00	0.00
60,000	26.28	9.28	17.98	15.54	30.84	0.00	0.00	0.00	0.00	0.00

active budget. In these examples, the maximal active budgets are 51,706.19, 21,086.23, and 50,196.49 in FC, OC, and PC models, so π^*_{FC}, π^*_{OC}, and π^*_{PC} in Table 16.2 do not change when B are larger than the maximal active budgets, respectively.

(3) The relative profit difference between FC and PC decreases as the value of B increases. When the budget B becomes larger, the total order $x^*_{i,\mathrm{PC}} + y^*_{i,\mathrm{PC}}$ will be close to $\tilde{x}_{i,\mathrm{PC}} + \tilde{y}_{i,\mathrm{PC}} = \max\left\{F_i^{-1}\left(\frac{p_i - v_i - w_i}{p_i - w_i}\right), F_i^{-1}\left(\frac{p_i - c_i}{p_i - s_i}\right)\right\}$, $i = 1, \ldots, n$, and the optimal order $x^*_{i,\mathrm{FC}}$ of fixed-price contract will be close to $\tilde{x}_{i,\mathrm{FC}} = F_i^{-1}\left(\frac{p_i - c_i}{p_i - s_i}\right)$, $i = 1, \ldots, n$. When the budget B becomes smaller, the difference between $x^*_{i,\mathrm{PC}} + y^*_{i,\mathrm{PC}}$ and $x^*_{i,\mathrm{FC}}$ become smaller, then the relative profit difference between FC and PC also becomes smaller.

(4) The relative profit difference between OC and PC decreases with the decreasing of B. When the budget B becomes smaller, $\Delta B^{\mathrm{O}}_{\mathrm{PC}}$ becomes larger, i.e., most of the budget will be spent on the option contracts, since option contracts occupy small unit procurement costs (i.e., $v_i < c_i$, $i = 1, \ldots, n$). Thus, the result of PC is close to that of OC as B decreases.

According to Tables 16.3 and 16.4, we know that: (1) The optimal order quantity of fixed-price contract in the portfolio contract increases as B increases, and it will become zero when B is small enough; (2) The optimal order quantity of option contract in the portfolio contract initially increases and then decreases with the increasing of B, and the turning point is $x^*_{i,\mathrm{PC}} > 0$, which is indicated in bold in

Table 16.5 Statistical comparison of the three different procurement models

Problem size n		10		50		100	
Relative profit difference (%)		$\Delta\pi_{\text{FC}}^{\text{PC}}$	$\Delta\pi_{\text{OC}}^{\text{PC}}$	$\Delta\pi_{\text{FC}}^{\text{PC}}$	$\Delta\pi_{\text{OC}}^{\text{PC}}$	$\Delta\pi_{\text{FC}}^{\text{PC}}$	$\Delta\pi_{\text{OC}}^{\text{PC}}$
Mean		21.86	25.35	22.09	25.64	21.89	25.82
Std. Dev.		3.47	3.54	1.61	1.56	1.17	1.07
95% C.I.	Lower	21.17	24.65	21.77	25.33	21.65	25.60
	Upper	22.55	26.06	22.41	25.95	22.12	26.03

Table 16.4. It will reach the minimal value $F_i^{-1}\left(\frac{p_i-w_i-v_i}{p_i-w_i}\right) - F_i^{-1}\left(\frac{w_i+v_i-c_i}{w_i-s_i}\right) > 0$ for the case of $\frac{p_i-w_i-v_i}{p_i-w_i} > \frac{p_i-c_i}{p_i-s_i}$ as B becomes large enough, and it will be zero for the case of $\frac{p_i-w_i-v_i}{p_i-w_i} < \frac{p_i-c_i}{p_i-s_i}$ as B becomes large enough.

From our theoretical results and the above observations, we come to the following insights: (1) managers should attempt to find FC strategy when the available budget is large and PC strategy is not available; (2) OC strategy should be paid more attention to when the available budget is too small and PC strategy cannot be used.

16.4.3 Strategies Comparison

In this section, the three procurement strategies, i.e., FC, OC, and PC, are compared by using randomly generated problems. In these examples, demands of all products are all normally distributed, and the total budget is $B = 3{,}500 \times n$. Let μ_i, σ_i, $i = 1,\ldots,n$, are parameters of the mean and standard deviation of the normal demand. We use the notation $x \sim U(\alpha,\beta)$ to denote that x is uniformly generated over $[\alpha,\beta]$. The problem parameters are generated as follows: $\mu_i \sim U(101,110)$, $\sigma_i \sim U(21,30)$, $p_i \sim (91,100)$, $c_i \sim U(41,50)$, $s_i \sim U(11,20)$, $w_i \sim U(41,50)$, $v_i \sim U(11,20)$, $i = 1,\ldots,n$. Note that the generated parameters satisfy the assumptions made in this chapter.

In this numerical study, we set $n = 10$, 50, 100, respectively. For each problem size n, 100 test instances are randomly generated. The statistical results of relative profit difference $\Delta\pi_{\text{FC}}^{\text{PC}}$ and $\Delta\pi_{\text{OC}}^{\text{PC}}$, are reported in Table 16.5, and the statistical results of computation time and number of iterations for searching λ^* in the portfolio contract model, are reported in Table 16.6. In these tables, 95% C.I. stands for 95% confidence interval.

From Table 16.5, we verify that the portfolio contract model outperforms the fixed-price contract and option contract models. This suggests that the retailer should pay more attention to portfolio contract when managing multi-product newsvendor problem with budget constraint if portfolio contracts are available. Additionally, the relationship between the two pure contract models depends on the problem parameters, e.g., in the 100 test instances for the case of $n = 10$, 34 option contract models outperform fixed-price contract models.

Table 16.6 Computation times and number of iterations of the solution method

		Computation time			Number of iterations		
Problem size n		10	50	100	10	50	100
Mean		10.92	29.17	52.27	30.10	31.81	32.88
Std. Dev.		1.91	2.35	3.90	3.08	2.66	2.61
95% C.I.	Lower	10.54	28.71	51.50	29.49	31.28	32.36
	Upper	11.30	29.64	53.04	30.71	32.34	33.40

According to Table 16.6, we know that our solution method can solve the problems quickly in limited iterations. The standard deviations of number of iterations and computation times are quite low in Table 16.6, reflecting the fact that our solution method is quite effective and robust. Robustness of our method should be attributed to the effectiveness of binary search procedure.

16.5 Conclusion

In this chapter, we investigate a portfolio approach to multi-product newsvendor problem with budget constraint, in which the procurement strategy for the newsvendor products is designed as portfolio contract. By establishing the structural properties of optimal solution, we develop an efficient solution method for the studied problem. The proposed algorithm has two main advantages: (1) it has linear computation complexity; (2) it is applicable to general continuous demand distribution.

In comparison with fixed-price contract and option contract models, the portfolio contract model generates significant improvement when managing multi-product newsvendor problem with budget constraint. Through sensitivity analysis, we come to the following insights: (1) The performance difference between fixed-price contract and portfolio contract models will become smaller as the available budget increases; (2) the performance gap between option contract and portfolio contract models increases when the available budget becomes larger. These insights suggest that managers with large budgets should pay more attention to fixed-price contract if the portfolio contract is not available, and that managers with small budgets should attempt to seek an option contract if the portfolio contract cannot be used.

There are several ways to extend this research. At first, this work can be directly extended to consider the scenario where different procurement strategies are available for different products. Secondly, another area of the future research is the consideration of an environment with supply uncertainty in sourcing and to investigate the effect of portfolio contract on managing supply uncertainty. Thirdly, one extension of this chapter might be to study portfolio contract model with demand updating in multi-stage settings. In addition, the demands for the multiple products are independent of one another in our study, an interesting and challenging extension is to consider the model in which the multiple products are substituted to some extent and thus the respective demands are correlated. Finally, it will be a significant issue to consider the model with the horizontal and/or vertical competition in the supply chain, and some topics of this extension have been investigated in a working paper of ours.

Acknowledgements The authors would like to thank the reviewers for their comments, which improved the manuscript. This work is supported by National Natural Science Foundation of China (Grants No. 70801065 and 71171206), National Natural Science Foundation of China major program (Grant No. 71090401/71090400), and the Fundamental Research Funds for the Central Universities of China.

Appendix:

Proof of Proposition 2. To prove this proposition, we consider two cases, respectively:

(1) $\lambda > \min\left\{\frac{p_i}{c_i}-1, \frac{p_i-w_i}{v_i}-1\right\}$, (2) $\lambda \le \min\left\{\frac{p_i}{c_i}-1, \frac{p_i-w_i}{v_i}-1\right\}$.

Case (1): The condition $\lambda > \min\left\{\frac{p_i}{c_i}-1, \frac{p_i-w_i}{v_i}-1\right\}$ holds only if $p_i - c_i < \lambda c_i$ or $p_i - w_i - v_i < \lambda v_i$.

If $p_i - c_i < \lambda c_i$, then we have $(p_i - s_i)F_i(x_i) < \alpha_i$ from (16.5). $(p_i - s_i)F_i(x_i) < \alpha_i$ and $\alpha_i x_i = 0$ implies $\tilde{x}_i^\lambda = 0$.

If $p_i - w_i - v_i < \lambda v_i$, then we have $(p_i - w_i)F_i(y_i) < \beta_i$ from (16.6). $(p_i - w_i)F_i(y_i) < \beta_i$ and $\beta_i y_i = 0$ implies $\tilde{y}_i^\lambda = 0$.

If $p_i - c_i < \lambda c_i$ and $p_i - w_i - v_i \ge \lambda v_i$, substituting $\tilde{x}_i^\lambda = 0$ into (16.6), we have $(p_i - w_i)F_i(y_i) \ge \beta_i$. $(p_i - w_i)F_i(y_i) \ge \beta_i$ and $\beta_i y_i = 0$ implies $\tilde{\beta}_i^\lambda = 0$. Then we have $\tilde{y}_i^\lambda = F_i^{-1}\left(\frac{p_i-w_i-(1+\lambda)v_i}{p_i-w_i}\right)$ from (16.6).

If $p_i - w_i - v_i < \lambda v_i$ and $p_i - c_i \ge \lambda c_i$, substituting $\tilde{y}_i^\lambda = 0$ into (16.5), we have $(p_i - s_i)F_i(x_i) \ge \alpha_i$. $(p_i - s_i)F_i(x_i) \ge \alpha_i$ and $\alpha_i x_i = 0$ implies $\tilde{\alpha}_i^\lambda = 0$. Then we have $\tilde{x}_i^\lambda = F_i^{-1}\left(\frac{p_i-(1+\lambda)c_i}{p_i-s_i}\right)$ from (16.5).

Thus, $\tilde{x}_i^\lambda = F_i^{-1}\left(\left(\frac{p_i-(1+\lambda)c_i}{p_i-s_i}\right)^+\right)$, and $\tilde{y}_i^\lambda = F_i^{-1}\left(\left(\frac{p_i-w_i-(1+\lambda)v_i}{p_i-w_i}\right)^+\right)$ if $\lambda > \min\left\{\frac{p_i}{c_i}-1, \frac{p_i-w_i}{v_i}-1\right\}$.

Case (2): According to (16.6), we have $x_i + y_i = F_i^{-1}\left(\frac{p_i-w_i-(1+\lambda)v_i+\beta_i}{p_i-w_i}\right)$. Substituting it into (16.5), we have $x_i = F_i^{-1}\left(\frac{w_i-(1+\lambda)(c_i-v_i)+\alpha_i-\beta_i}{w_i-s_i}\right)$. The condition $\lambda \le \min\left\{\frac{p_i}{c_i}-1, \frac{p_i-w_i}{v_i}-1\right\}$ implies $p_i - c_i \ge \lambda c_i$ and $p_i - w_i - v_i \ge \lambda v_i$.

In this case, we consider three subcases:

(2.1) $$\frac{p_i-(1+\lambda)c_i}{p_i-s_i} < \frac{p_i-w_i-(1+\lambda)v_i}{p_i-w_i},$$

(2.2) $$\frac{p_i-w_i-(1+\lambda)v_i}{p_i-w_i} < \frac{p_i-(1+\lambda)c_i}{p_i-s_i},$$

(2.3) $$\frac{p_i-w_i-(1+\lambda)v_i}{p_i-w_i} = \frac{p_i-(1+\lambda)c_i}{p_i-s_i}.$$

Subcase (2.1): In this case, we have $\frac{w_i-(1+\lambda)(c_i-v_i)}{w_i-s_i} < \frac{p_i-(1+\lambda)c_i}{p_i-s_i} < \frac{p_i-w_i-(1+\lambda)v_i}{p_i-w_i}$. By combining (16.5) and (16.6), we have $w_i-(1+\lambda)(c_i-v_i)-\beta_i = (w_i-s_i)F_i(x_i)-\alpha_i$.

If $w_i-(1+\lambda)(c_i-v_i)<0$, then we have $(w_i-s_i)F_i(x_i)<\alpha_i$. $(w_i-s_i)F_i(x_i)<\alpha_i$ and $\alpha_i x_i=0$ implies $\tilde{x}_i^\lambda=0$. Substituting $\tilde{x}_i^\lambda=0$ into (16.6), we have $(p_i-w_i)F_i(y_i)>\beta_i$. $(p_i-w_i)F_i(y_i)>\beta_i$ and $\beta_i y_i=0$ implies $\tilde{\beta}_i^\lambda=0$. Then we have $\tilde{y}_i^\lambda = F_i^{-1}\left(\frac{p_i-w_i-(1+\lambda)v_i}{p_i-w_i}\right)$ from (16.6).

If $w_i-(1+\lambda)(c_i-v_i)\geq 0$, then $w_i-(1+\lambda)(c_i-v_i)=(w_i-s_i)F_i(x_i)-\alpha_i+\beta_i\geq 0$. Since $p_i-w_i-(1+\lambda)v_i>0$, there must be $x_i+y_i>0$. According to $\alpha_i x_i+\beta_i y_i=0$, we know $\alpha_i\beta_i=0$. If $\beta_i=0$, then $(w_i-s_i)F_i(x_i)-\alpha_i\geq 0$ and $\alpha_i x_i=0$ implies $\alpha_i=0$. If $\alpha_i=0$, then $x_i+y_i>x_i$ implies $y_i>0$, and then $\beta_i=0$. Thus $\tilde{\alpha}_i^\lambda=\tilde{\beta}_i^\lambda=0$, and

$$\tilde{x}_i^\lambda = F_i^{-1}\left(\frac{w_i-(1+\lambda)(c_i-v_i)}{w_i-s_i}\right),$$

$$\tilde{y}_i^\lambda = F_i^{-1}\left(\frac{p_i-w_i-(1+\lambda)v_i}{p_i-w_i}\right) - F_i^{-1}\left(\frac{w_i-(1+\lambda)(c_i-v_i)}{w_i-s_i}\right).$$

These results in subcase (2.1) can be rewritten as:

$$\tilde{x}_i^\lambda = F_i^{-1}\left(\left(\frac{w_i-(1+\lambda)(c_i-v_i)}{w_i-s_i}\right)^+\right),$$

$$\tilde{y}_i^\lambda = F_i^{-1}\left(\frac{p_i-w_i-(1+\lambda)v_i}{p_i-w_i}\right) - F_i^{-1}\left(\left(\frac{w_i-(1+\lambda)(c_i-v_i)}{w_i-s_i}\right)^+\right),$$

if $\lambda \leq \min\left\{\frac{p_i}{c_i}-1, \frac{p_i-w_i}{v_i}-1\right\}$ and $\frac{p_i-(1+\lambda)c_i}{p_i-s_i} < \frac{p_i-w_i-(1+\lambda)v_i}{p_i-w_i}$.

Subcase (2.2): In this case, we have $\frac{p_i-w_i-(1+\lambda)v_i}{p_i-w_i} < \frac{p_i-(1+\lambda)c_i}{p_i-s_i} < \frac{w_i-(1+\lambda)(c_i-v_i)}{w_i-s_i}$. $x_i+y_i\geq x_i$ requires $\tilde{\beta}_i^\lambda>0$, and, hence, $\tilde{y}_i^\lambda=0$. Since $\tilde{x}_i^\lambda = F_i^{-1}\left(\frac{p_i-w_i-(1+\lambda)v_i+\tilde{\beta}_i^\lambda}{p_i-w_i}\right)$ >0, $\alpha_i x_i=0$ implies $\tilde{\alpha}_i^\lambda=0$. $\frac{w_i-(1+\lambda)(c_i-v_i)-\tilde{\beta}_i^\lambda}{w_i-s_i} = \frac{p_i-w_i-(1+\lambda)v_i+\tilde{\beta}_i^\lambda}{p_i-w_i}$ implies

$$\tilde{\beta}_i^\lambda = \frac{(p_i - w_i)(w_i - s_i)}{p_i - s_i}\left(\frac{w_i - (1+\lambda)(c_i - v_i)}{w_i - s_i} - \frac{p_i - w_i - (1+\lambda)v_i}{p_i - w_i}\right).$$

Substituting it into $\tilde{x}_i^\lambda$, we have $\tilde{x}_i^\lambda = F_i^{-1}\left(\frac{p_i-(1+\lambda)c_i}{p_i-s_i}\right)$.

Subcase (2.3): In this case, we have $\frac{p_i-w_i-(1+\lambda)v_i}{p_i-w_i} = \frac{p_i-(1+\lambda)c_i}{p_i-s_i} = \frac{w_i-(1+\lambda)(c_i-v_i)}{w_i-s_i}$. If $\alpha_i > 0$ and $\beta_i = 0$, then $x_i + y_i < x_i$, it is in contradiction with $x_i + y_i \geq x_i$. If $\alpha_i = 0$ and $\beta_i > 0$, then $x_i < x_i + y_i$, and, hence, $y_i > 0$; It is in contradiction with $\beta_i y_i = 0$. If $\alpha_i > 0$ and $\beta_i > 0$, then $x_i + y_i > 0$, and, hence, $\alpha_i x_i + \beta_i y_i \neq 0$, which violates the slackness condition. Thus, there must be $\tilde{\alpha}_i^\lambda = \tilde{\beta}_i^\lambda = 0$, then $\tilde{x}_i^\lambda + \tilde{y}_i^\lambda = \tilde{x}_i^\lambda$, and $\tilde{y}_i^\lambda = 0$, $\tilde{x}_i^\lambda = F_i^{-1}\left(\frac{p_i-(1+\lambda)c_i}{p_i-s_i}\right)$.

Thus, the results in subcases (2.2) and (2.3) are both $\tilde{x}_i^\lambda = F_i^{-1}\left(\frac{p_i-(1+\lambda)c_i}{p_i-s_i}\right)$, and $\tilde{y}_i^\lambda = 0$, if $\lambda \leq \min\left\{\frac{p_i}{c_i} - 1, \frac{p_i-w_i}{v_i} - 1\right\}$ and $\frac{p_i-w_i-(1+\lambda)v_i}{p_i-w_i} \leq \frac{p_i-(1+\lambda)c_i}{p_i-s_i}$.

In summary, all these results in the three subcases can be generalized as the equations in Proposition 2. □

Proof of Proposition 4. To prove this proposition, we first cite the following results from the proof of Proposition 2:

(a) If $\lambda > \min\{\frac{p_i}{c_i} - 1, \frac{p_i-w_i}{v_i} - 1\}$, then $\tilde{x}_i^\lambda = F_i^{-1}((\frac{p_i-(1+\lambda)c_i}{p_i-s_i})^+)$, and $\tilde{y}_i^\lambda = F_i^{-1}((\frac{p_i-w_i-(1+\lambda)v_i}{p_i-w_i})^+)$;
(b) If $\lambda \leq \min\left\{\frac{p_i}{c_i} - 1, \frac{p_i-w_i}{v_i} - 1\right\}$ and $\frac{p_i-(1+\lambda)c_i}{p_i-s_i} < \frac{p_i-w_i-(1+\lambda)v_i}{p_i-w_i}$, then

$$\tilde{x}_i^\lambda = F_i^{-1}\left(\left(\frac{w_i-(1+\lambda)(c_i-v_i)}{w_i-s_i}\right)^+\right), \quad \tilde{y}_i^\lambda = F_i^{-1}\left(\frac{p_i-w_i-(1+\lambda)v_i}{p_i-w_i}\right)$$
$$-F_i^{-1}\left(\left(\frac{w_i-(1+\lambda)(c_i-v_i)}{w_i-s_i}\right)^+\right);$$

(c) If $\lambda \leq \min\left\{\frac{p_i}{c_i} - 1, \frac{p_i-w_i}{v_i} - 1\right\}$ and $\frac{p_i-w_i-(1+\lambda)v_i}{p_i-w_i} \leq \frac{p_i-(1+\lambda)c_i}{p_i-s_i}$, then

$$\tilde{x}_i^\lambda = F_i^{-1}\left(\frac{p_i-(1+\lambda)c_i}{p_i-s_i}\right), \quad \tilde{y}_i^\lambda = 0.$$

Under case (a) or (c), the result in this proposition is obvious. Under case (b), we have $\tilde{x}_i^\lambda = F_i^{-1}((\frac{w_i-(1+\lambda)(c_i-v_i)}{w_i-s_i})^+)$ and $\tilde{x}_i^\lambda + \tilde{y}_i^\lambda = F_i^{-1}(\frac{p_i-w_i-(1+\lambda)v_i}{p_i-w_i})$. Thus $c_i\tilde{x}_i^\lambda + v_i\tilde{y}_i^\lambda = (c_i - v_i)\tilde{x}_i^\lambda + v_i(\tilde{x}_i^\lambda + \tilde{y}_i^\lambda)$ is nonincreasing in λ. □

References

Abdel-Malek, L., Montanari, R., & Morales, L. C. (2004). Exact, approximate, and generic iterative models for the multi-product newsboy with budget constraint. *International Journal of Production Economics, 91*, 189–198.

Abdel-Malek, L., & Montanari, R. (2005). An analysis of the multi-product newsboy problem with a budget constraint. *International Journal of Production Economics, 97*, 296–307.

Aggarwal, P., & Ganeshan, R. (2007). Using risk-management tools on B2Bs: an exploratory investigation. *International Journal of Production Economics, 108*, 2–7.

Barnes-Schuster, D., Bassok, Y., & Anupindi, R. (2002). Coordination and flexibility in supply contracts with options. *Manufacturing and Service Operations Management, 4(3)*, 171–207.

Boeckem, S. & Schiller, U. (2008). Option contracts in supply chains. *Journal of Economics and Management Strategy, 17*, 219–245.

Cachon, G. (2003). Supply chain coordination with contracts. In S. Graves, T. de Kok (Eds.), *The handbook of operations research and management science: Supply chain management*. Amsterdam: Kluwer.

Carbone, J. (2001). HP buyers get hands on design. *Purchasing.com*, 19 July 2001.

Cohen, M., & Agrawal, N. (1999). An analytical comparison of long and short term contracts. *IIE Transaction, 31(8)*, 783–796.

Marquez, A. C., & Blanchar, C. (2004). The procurement of strategic parts. Analysis of a portfolio of contracts with suppliers using a system dynamics simulation model. *International Journal of Production Economics, 88*, 29–49.

Erlebacher, S. J. (2000). Optimal and heuristic solutions for the multi-item newsvendor problem with a single capacity constraint. *Production and Operations Management, 9(3)*, 303–318.

Fu, Q., Lee, C. Y., & Teo, C. P. (2010). Procurement management using option contracts: random spot price and the portfolio effect. *IIE Transactions, 42(11)*, 793–811.

Gomez_Padilla, A., & Mishina, T. (2009). Supply contract with options. *International Journal of Production Economics*, 122(1), 312–318.

Hadley, G. & Whitin, T. (1963). *Analysis of inventory systems*. Prentice-Hall: Englewood Cliffs.

Khouja, M. (1999). The single-period (news-vendor) problem: literature review and suggestions for future research. *Omega, 27*, 537–553.

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

Lau, H. S., & Lau, A. H. L. (1996). The newsstand problem: a capacitated multiple-product single-period inventory problem. *European Journal of Operational Research, 94*, 29–42.

Martínez de Albéniz, V., & Simchi-Levi, D. (2005). A portfolio approach to procurement contracts. *Production and Operations Management, 14(1)*, 90–114.

Moon, I., & Silver, E. (2000). The multi-item newsvendor problem with a budget constraint and fixed ordering costs. *Journal of the Operational Research Society, 51*, 602–608.

Vairaktarakis, G.L. (2000). Robust multi-item newsboy models with a budget constraint. *International Journal of Production Economics, 66*, 213–226.

Wang, X., & Liu L. (2007). Coordination in a retailer-led supply chain through option contract. *International Journal of Production Economics, 110*, 115–127.

Wang, Q., & Tsao D. (2006) Supply contract with bidirectional options: the buyer's perspective. *International Journal of Production Economics, 101*, 30–52.

Wu, D. J., & Kleindorfer, P. R. (2005). Competitive options, supply contracting, and electronic markets. *Management Science, 51(3)*, 452–466.

Wu, D. J., Kleindorfer, P., & Zhang, J. E. (2002). Optimal bidding and contracting strategies for capital-intensive goods. *European Journal of Operational Research, 137(3)*, 657–676.

Zhang, B., & Du, S.F. (2010). Multi-product newsboy problem with limited capacity and outsourcing. *European Journal of Operational Research, 202*(1), 107–113.

Zhang, B., & Hua, Z. (2008). A unified method for a class of convex separable nonlinear knapsack problem. *European Journal of Operational Research, 191(1)*, 1–6.

Zhang, B., Xu, X., & Hua, Z. (2009). A binary solution method for the multi-product newsboy problem with budget constraint. *International Journal of Production Economics, 117*, 136–141.

Zhang, B. (2011). Optimal policy for a mixed production system with multiple OEM and OBM products. *International Journal of Production Economics, 130*(1), 27–32

Zhang, B. (2012). Multi-tier binary solution method for multi-product newsvendor problem with multiple constraints. *European Journal of Operational Research*, 218(2), 426–434.

Zhao, Y., Wang, S., Cheng, T. C. E., Yang, X., & Huang, Z. (2010). Coordination of supply chains by option contracts: a cooperative game theory approach. *European Journal of Operational Research, 207*(2), 668–675.

Index

T.-M. Choi (ed.), *Handbook of Newsvendor Problems*, International Series in Operations Research & Management Science 176, DOI 10.1007/978-1-4614-3600-3,