ΠΑΝΕΠΙΣΤΗΜΙΟ ΘΕΣΣΑΛΙΑΣ

ΠΟΛΥΤΕΧΝΙΚΗ ΣΧΟΛΗ

ΤΜΗΜΑ ΜΗΧΑΝΟΛΟΓΩΝ ΜΗΧΑΝΙΚΩΝ ΒΙΟΜΗΧΑΝΙΑΣ

Διατριβή

ΠΟΣΟΤΙΚΗ ΑΝΑΛΥΣΗ ΤΗΣ ΕΠΙΔΡΑΣΗΣ ΤΗΣ ΕΛΛΕΙΨΗΣ ΠΡΟΪΟΝΤΩΝ ΣΤΗ ΔΙΑΧΕΙΡΙΣΗ ΑΠΟΘΕΜΑΤΩΝ

υπό

ΙΣΙΔΩΡΟΥ ΤΣΙΚΗ

Διπλωματούχου Μηχανολόγου Μηχανικού Βιομηχανίας Πανεπιστημίου Θεσσαλίας, 2000
Μ.Δ.Ε. Μηχανολόγου Μηχανικού Βιομηχανίας Πανεπιστημίου Θεσσαλίας, 2003

Υπεβλήθη για την εκπλήρωση μέρους των

απαιτήσεων για την απόκτηση του

Διδακτορικού Διπλώματος

2007

Εγκρίθηκε από τα Μέλη της Επταμελούς Εξεταστικής Επιτροπής:

Πρώτος Εξεταστής
(Επιβλέπων)

Δρ. Γεώργιος Λυμπερόπουλος
Αναπληρωτής Καθηγητής Τμήματος Μηχανολόγων Μηχανικών Βιομηχανίας
Πανεπιστημίου Θεσσαλίας

Δεύτερος Εξεταστής

Δρ. Αθανάσιος Ζηλιασκόπουλος
Αναπληρωτής Καθηγητής Τμήματος Μηχανολόγων Μηχανικών Βιομηχανίας
Πανεπιστημίου Θεσσαλίας

Τρίτος Εξεταστής

Δρ. Γεώργιος Ταγαράς
Καθηγητής Τμήματος Μηχανολόγων Μηχανικών Αριστοτελείου
Πανεπιστημίου Θεσσαλονίκης

Τέταρτος Εξεταστής

Δρ. Χρυσολέων Παπαδόπουλος
Καθηγητής Τμήματος Οικονομικών Επιστημών Αριστοτελείου Πανεπιστημίου
Θεσσαλονίκης

Πέμπτος Εξεταστής

Δρ. Γρηγόριος Πραστάκος
Καθηγητής Τμήματος Διοικητικής Επιστήμης και Τεχνολογίας Οικονομικού
Πανεπιστημίου Αθηνών

Έκτος Εξεταστής

Δρ. Απόστολος Μπουρνέτας
Αναπληρωτής Καθηγητής Τμήματος Μαθηματικών Εθνικού &
Καποδιστριακού Πανεπιστημίου Αθηνών

Έβδομος Εξεταστής

Δρ. Γεώργιος Ιωάννου
Αναπληρωτής Καθηγητής Τμήματος Διοικητικής Επιστήμης και Τεχνολογίας
Οικονομικού Πανεπιστημίου Αθηνών

Ευχαριστίες

Κατ’ αρχάς, θέλω να ευχαριστήσω τον επιβλέποντα της διατριβής μου, Αναπληρωτή Καθηγητή κ. Γιώργο Λυμπερόπουλο, για την πολύτιμη βοήθεια και καθοδήγησή του κατά τη διάρκεια της διδακτορικής μου διατριβής. Επίσης, είμαι ευγνώμων στα υπόλοιπα μέλη της εξεταστικής επιτροπής της διατριβής μου, Καθηγητές κκ. Ζηλιασκόπουλο, Ταγαρά, Παπαδόπουλο, Πραστάκο, Μπουρνέτα και Ιωάννου, για την προσεκτική ανάγνωση της εργασίας μου και για τις χρήσιμες υποδείξεις τους. Ιδιαίτερες ευχαριστίες οφείλω στον Δρ. Γιώργο Κοζανίδη για την πολύτιμη βοήθειά του στον προγραμματισμό των προτύπων παιγνίων του Κεφαλαίου 3 της διατριβής μου, καθώς και για την φιλία του. Επίσης, ευχαριστώ τη σύζυγό μου Ελένη Κατσαρού για την κατανόηση, συμπαράσταση και υπομονή της, καθώς και τον αδελφό μου, Τάσο Τσική, για την συμπαράστασή του. Πάνω απ’ όλα, είμαι ευγνώμων στους γονείς μου, Γιάννη και Μαρία Τσική, για την ολόψυχη αγάπη και υποστήριξή τους όλα αυτά τα χρόνια. Αφιερώνω αυτή την διατριβή στην μητέρα μου και στην μνήμη του πατέρα μου. Και οι δύο έκαναν αμέτρητες θυσίες για να μπορέσω να την ολοκληρώσω. Η εκπόνηση αυτής της διατριβής υποστηρίχθηκε από την Πράξη «Ηράκλειτος: Υποτροφίες Έρευνας με Προτεραιότητα στην Βασική Έρευνα» του Επιχειρησιακού Προγράμματος «Εκπαίδευση και Αρχική Επαγγελματική Κατάρτιση» (ΕΠΕΑΕΚ) ΙΙ του Υπουργείου Εθνικής Παιδείας και Θρησκευμάτων που συγχρηματοδοτήθηκε από το Ευρωπαϊκό Κοινοτικό Ταμείο, το Ευρωπαϊκό Ταμείο Περιφερειακής Ανάπτυξης και το Ελληνικό Δημόσιο.

Ισίδωρος Τσική

ΠΟΣΟΤΙΚΗ ΑΝΑΛΥΣΗ ΤΗΣ ΕΠΙΔΡΑΣΗΣ ΤΗΣ ΕΛΛΕΙΨΗΣ ΠΡΟΪΟΝΤΩΝ ΣΤΗ ΔΙΑΧΕΙΡΙΣΗ ΑΠΟΘΕΜΑΤΩΝ

ΙΣΙΔΩΡΟΣ ΤΣΙΚΗΣ

Πανεπιστήμιο Θεσσαλίας, Πολυτεχνική Σχολή, Τμήμα Μηχανολόγων Μηχανικών Βιομηχανίας, 2007

Επιβλέπων Καθηγητής: Δρ. Γεώργιος Λυμπερόπουλος, Αναπληρωτής Καθηγητής Διοίκησης Παραγωγής

Περίληψη

Έλλειψη προϊόντος (stockout) σημειώνεται κάθε φορά που ένας πελάτης παραγγέλνει ένα προϊόν από έναν προμηθευτή και ο προμηθευτής δεν μπορεί να το παραδώσει άμεσα επειδή δεν το έχει στο απόθεμά του. Μια έλλειψη μπορεί να επιφέρει άμεσα ή/και έμμεσα κόστη στον προμηθευτή. Τα έμμεσα κόστη σχετίζονται με την απώλεια καλής πίστης των πελατών προς τον προμηθευτή μετά από μια έλλειψη και μπορεί να οδηγήσει σε μείωση της μελλοντικής ζήτησης και του μεριδίου αγοράς του προμηθευτή. Το κίνητρο αυτής της διατριβής προέρχεται από την ανάγκη της ποσοτικοποίησης των έμμεσων κοστών που προκαλούνται από τις ελλείψεις, ενός ζητήματος που δεν έχει αντιμετωπισθεί ικανοποιητικά στη βιβλιογραφία. Η διατριβή χωρίζεται σε τρία μέρη που διαφέρουν μεταξύ τους ως προς την οπτική γωνία εξέτασης του προβλήματος.

Στο πρώτο μέρος εξετάζουμε το κλασικό πρότυπο αποθεμάτων Βέλτιστης Ποσότητας Παραγγελίας (Economic Order Quantity ή EOQ) με ελλείψεις που κοστολογούνται με έναν σταθερό συντελεστή κόστους έλλειψης προϊόντων «*b*». Ο συντελεστής αυτός αντικατοπτρίζει την απώλεια καλής πίστης – και άρα και της μελλοντικής ζήτησης – των πελατών προς τον προμηθευτή λόγω των ελλείψεων. Το ερώτημα που θέτουμε είναι πώς θα μπορούσαμε να εκτιμήσουμε την τιμή του *b*; Για να απαντήσουμε στο ερώτημα αυτό, συνδέουμε το κλασσικό πρότυπο αποθεμάτων EOQ με ελλείψεις με ένα αντίστοιχο πρότυπο διαχείρισης αποθεμάτων με «διαταραγμένη» ζήτηση, στο οποίο δεν υπάρχει κόστος έλλειψης προϊόντων, αλλά η μακροπρόθεσμη μέση ζήτηση είναι μια αύξουσα συνάρτηση του επιπέδου εξυπηρέτησης πελατών. Η χρήση του προτύπου διαταραγμένης ζήτησης αντικαθιστά το δυσχερές έργο υπολογισμού του *b* στο κλασσικό πρότυπο EOQ με το ευκολότερο έργο υπολογισμού των παραμέτρων της συνάρτησης της διαταραγμένης ζήτησης.

Στο δεύτερο μέρος αναπτύσσουμε ένα πρότυπο τύπου «εφημεριδοπώλη» δύο προμηθευτών που ανταγωνίζονται να πουλήσουν τον ίδιο τύπο προϊόντων σε ένα πελάτη, επαναλαμβανόμενα, σε

διακριτές περιόδους, για άπειρο χρονικό ορίζοντα. Στην αρχή κάθε περιόδου, κάθε προμηθευτής παραγγέλνει έναν αριθμό προϊόντων που του παραδίδονται άμεσα. Σε κάθε περίοδο, ο πελάτης επιλέγει τυχαία έναν από τους δύο προμηθευτές και του ζητάει έναν τυχαίο αριθμό προϊόντων. Η πιθανότητα με την οποία επιλέγεται ένας προμηθευτής εξαρτάται από το λεγόμενο «επίπεδο αξιοπιστίας» αυτού του προμηθευτή, το οποίο αντικατοπτρίζει την εκτίμηση που έχει ο πελάτης για τη σχετική αξιοπιστία του προμηθευτή βάσει της ιστορίας εξυπηρέτησης – μετρούμενης με όρους διαθεσιμότητας προϊόντων – που έχουν παράσχει και οι δύο προμηθευτές στον πελάτη στο παρελθόν. Τα επίπεδα αξιοπιστίας των προμηθευτών αλλάζουν δυναμικά ανάλογα με την ποιότητα εξυπηρέτησης – καλής ή κακής – που ο πελάτης λαμβάνει σε κάθε περίοδο. Μορφοποιούμε το πρόβλημα εύρεσης βέλτιστων στάσιμων πολιτικών παραγγελίας για τους δύο προμηθευτές σε ισορροπία ως ένα στοχαστικό δυναμικό παίγνιο, και επιλύουμε αριθμητικά τις προκύπτουσες συνθήκες βελτιστότητας για διάφορες περιστάσεις του προβλήματος. Σε όλες τις περιστάσεις, προκύπτει ότι η βέλτιστη πολιτική παραγγελίας για κάθε προμηθευτή είναι μια πολιτική τύπου «παραγγελίας μέχρις ένα επίπεδο» (order-up-to). Στη συνέχεια, εξετάζουμε μια συγκεκριμένη περίπτωση του προτύπου, όπου κάθε προμηθευτής έχει μόνο δύο ακραία επίπεδα αξιοπιστίας, ένα υψηλό και ένα χαμηλό, τέτοια ώστε όταν βρίσκεται στο χαμηλό επίπεδο, δεν επιλέγεται ποτέ από τον πελάτη, ενώ όταν βρίσκεται στο υψηλό, πάντα επιλέγεται. Γι' αυτήν την περίπτωση υποθέτουμε ότι και οι δύο προμηθευτές χρησιμοποιούν μια πολιτική τύπου παραγγελίας μέχρις ένα επίπεδο, όπου το επίπεδο παραγγελίας εξαρτάται από το επίπεδο αξιοπιστίας. Μορφοποιούμε το πρόβλημα ως μια Μαρκοβιανή Διαδικασία Αποφάσεων με δύο κέντρα απόφασης. Για μια ειδική κατανομή της ζήτησης αποδεικνύουμε την ύπαρξη ενός μοναδικού σημείου ισορροπίας Nash, και επιλύουμε αριθμητικά τις προκύπτουσες συνθήκες βελτιστότητας στην κατάσταση ισορροπίας για να βρούμε τα βέλτιστα επίπεδα παραγγελίας και των δύο προμηθευτών.

Στο τρίτο και τελευταίο μέρος αναζητούμε εμπειρικά τεκμήρια ότι οι ελλείψεις προϊόντων πράγματι επηρεάζουν αρνητικά τη μελλοντική ζήτηση. Για αυτό το σκοπό, εξετάζουμε τη σύνδεση μεταξύ ελλείψεων, εξυπηρετήσεων πελατών, τρεχουσών πωλήσεων και μελλοντικής ζήτησης, πραγματοποιώντας εκτενή στατιστική ανάλυση ιστορικών δεδομένων πωλήσεων και παραδόσεων μιας πραγματικής χονδρεμπορικής εταιρείας για μια περίοδο τεσσάρων ετών. Το βασικό μας εύρημα είναι ότι οι ελλείψεις έχουν πράγματι αρνητική επίδραση στις τρέχουσες πωλήσεις και στη μελλοντική ζήτηση.

UNIVERSITY OF THESSALY

SCHOOL OF ENGINEERING

DEPARTMENT OF MECHANICAL & INDUSTRIAL ENGINEERING

Dissertation

QUANTITATIVE ANALYSIS OF THE IMPACT OF STOCKOUTS IN INVENTORY MANAGEMENT

By

ISIDOROS TSIKIS

Diploma in Mechanical & Industrial Engineering, University of Thessaly, 2000
M.Sc. in Mechanical & Industrial Engineering, University of Thessaly, 2003

Submitted in partial fulfillment of the

requirements for the degree of

Doctor of Philosophy

2007

The approval of this Ph.D. Dissertation by the Department of Mechanical and Industrial Engineering of the School of Engineering of the University of Thessaly does not imply acceptance of the writer's opinions. (Law 5343/32, article 202 par. 2).

Acknowledgements

First and foremost, I would like to thank my thesis supervisor, Associate Professor George Liberopoulos, for his valuable help and guidance throughout my doctoral work. I am also grateful to the readers of my thesis, Professors Ziliaskopoulos, Tagaras, Papadopoulos, Prastacos, Burnetas, and Ioannou, for carefully reading my work and making valuable suggestions. Due thanks are extended to Dr. George Kozanidis for his valuable help in programming the competitive game models in Chapter 3 of my thesis, as well as for his friendship. I would also like to thank my wife Eleni Katsarou for her understanding, support, and patience, as well as my brother, Tassos Tsikis, for his support. Most of all, I am grateful to my parents, Ioannis and Maria Tsikis, for their wholehearted love and support all these years. I dedicate this thesis to my mother and to the memory of my father. They both made innumerable sacrifices so that I could complete it. The work in this thesis was supported by the Action “Heraclitus: Research Scholarships with Priority in Basic Research” of the Operational Program for Education and Initial Vocational Training II of Greece’s Ministry of National Education and Religious Affairs, which was co-financed by the European Social Fund, the European Regional Development Fund, and Greece’s Public Sector.

Isidoros Tsikis

QUANTITATIVE ANALYSIS OF THE IMPACT OF STOCKOUTS IN INVENTORY MANAGEMENT

ISIDOROS TSIKIS

University of Thessaly, School of Engineering, Department of Mechanical & Industrial Engineering, 2007

Dissertation Supervisor: Dr. George Liberopoulos, Associate Professor of Production Management

Abstract

A stockout occurs whenever a customer requests an item from a supplier and the supplier can not deliver this item due to a temporary lack of stock. A stockout may incur direct or indirect costs to the supplier. The indirect costs are related to the loss of customer goodwill following a stockout which may lead to a decline in future demand and market share of a supplier, especially in a competitive market environment. This thesis is motivated by the need to quantify the indirect costs of stockouts, which has long been an unsatisfactorily resolved issue in the literature. It is divided into three parts that differ from each other in their perspective.

In the first part, we revisit the classical Economic Order Quantity (EOQ) model with backorders that are being penalized with a backorder penalty cost coefficient, b. This coefficient reflects the intangible effect of the future loss of customer goodwill – and therefore demand – following a stockout. We ask the question, what could b be? To answer this question, we infer the value of b in the EOQ model with penalized backorders by connecting this model to a perturbed demand model which assumes that there is no explicit backorder penalty cost, but that the long-run average demand rate is an increasing function of the customer service level. The perturbed demand model replaces the impracticable task of estimating b in the classical model with penalized backorders, with the more feasible task of estimating the parameters of the perturbed demand rate function.

In the second part, we develop a newsvendor-type model of two suppliers that compete to sell the same type of items to a customer, repetitively, in discrete periods, for an infinite time horizon. At the beginning of each period, each supplier orders a number of items that are delivered to him immediately. In each period, the customer randomly chooses one of the two suppliers and demands from him a random number of items. The probability of choosing a supplier depends on

the so-called “credibility level” of this supplier, which reflects the customer's estimate of the supplier's relative credibility based on the history of service – measured in terms of product availability – that both suppliers have provided to the customer in the past. The credibility levels of the suppliers change dynamically based on the quality of service – good or poor – that the customer receives in each period. We formulate the problem of finding optimal stationary ordering policies for both suppliers at equilibrium as a stochastic dynamic game, and we numerically solve the resulting optimality conditions for several instances of this problem. In all instances, the optimal ordering policy for each supplier turns out to be an order-up-to policy. Then, we restrict our attention to the case where each supplier has only two extreme credibility levels, a low and a high, such that, when in the low level, he is never chosen by the customer, and when in the high level, he is always chosen by the customer. For this case, we assume that each supplier uses a credibility level-dependent order-up-to policy. This leads to a Markov Decision Process with two decision makers. For a special demand distribution, we show that there exists a unique Nash equilibrium, and we numerically solve the resulting optimality conditions at equilibrium to find the optimal order-up-to levels of both suppliers.

In the third and last part, we seek empirical evidence that stockouts affect future demand. To this end, we study the linkage between stockouts, customer service, current sales, and future demand, by performing a thorough statistical analysis of historical customer order and delivery data of a tool wholesaler and distributor over a period of four years. Our main finding is that stockouts do have an adverse effect on current sales and future customer demand.

Contents

List of Tables

List of Figures

Chapter 1 Introduction

In this chapter, we provide some background information which supports the motivation behind this thesis. We also review the relevant literature, and we give a brief description of the three main parts of the thesis, which occupy Chapters 2-4, respectively.

1.1 Motivation and Background

Anyone who has taken or taught a course in inventory management is likely to have pondered at how to quantify the cost incurred by a stockout. A stockout occurs whenever a customer requests an item from a supplier and the supplier can not deliver this item due to a temporary lack of stock. Stockouts can be either internal or external to a firm. An internal stockout occurs when an order of a department within the organization is not filled. An external stockout occurs when the supplier does not fill a customer's order on time. Internal stockouts can result in lost production and delays. Stockouts may incur direct costs which may be analyzed into backorder costs and current profit losses. Backorder costs typically include extra costs for administration, price discounts or contractual penalties for late deliveries, expediting material handling and transportation, the potential interest on the profit tied up in the backorder, etc. The current profit loss is the potential profit of the sale, if the sale is lost.

Although the direct costs incurred by stockouts may be quite significant, what most researchers and practitioners understand as stockout costs are the indirect costs that are related to the loss of customer goodwill following a stockout which may lead to a temporary or permanent decline in future demand and market share of a supplier, especially in a competitive market environment. In the short run, the supplier's sales may fall short of demand when customers experience stockouts and choose not to backorder. In the long run, the supplier's demand itself may decline as customers who experience excessive stockouts shift permanently to more reliable sources.

The quantification of the indirect cost of stockouts has long been an unsatisfactorily resolved issue in the literature. Yet, most of the traditional approaches to determining optimal or simply good inventory control policies have been based on assuming a specific functional form for stockout costs. A commonly used assumption is that backorders are being penalized with a constant backorder penalty cost rate, which is often denoted by b. The difficulty in determining an appropriate value for b has prompted many researchers to replace the backorder penalty cost by a constraint on the customer service level. This may be more appealing to practitioners, but it only transposes the problem of estimating the appropriate cost of stockouts to one of determining the appropriate customer service level. An inquisitive student may still wonder why a 95% service level, which is often used as an example in textbooks, is better than a 94% or a 96% service level.

The difficulty in estimating b or its surrogate service level target lies in the fact that b is supposed to reflect the intangible effect of the loss of customer goodwill following a stockout. As Schwartz (1966) noted, however, the effect of the loss of goodwill should not be a penalty cost of the type considered in the classical inventory models. This is because the backorder penalty cost term in the objective function of such models is subtracted as though the firm incurs an expense at the time of the stockout. Yet, the effect of the loss of customer goodwill is incurred not at the time of the stockout incident, but at a later time, due to the customer's decision to alter his future demand. With this in mind, Schwartz (1966) modified the classical Economic Order Model (EOQ) with backorders by eliminating the explicit backorder penalty cost term from the objective function and by assuming that the long-run demand rate – and hence the long-run average reward – is an increasing function of the fill rate. Schwartz called the resulting model a "perturbed demand" (PD) model.

We agree with Schwartz that the PD approach to goodwill stockout penalties is more valid than the classical inventory theory approach, for two reasons. The first reason is the already stated difficulty in picking a good – let alone the best – value for the backorder (or stockout) coefficient or the customer service level in the classical approach. The second reason is that the classical approach has the following paradox embedded in it. It supposes that there is a backorder penalty cost which reflects the future loss of demand due to the loss of customer goodwill following stockouts, and yet it assumes that the demand is stationary. While we find that the PD approach to goodwill stockout penalties is more valid than the classical inventory theory approach, we are not sure if it is more practical than the classical approach. If it were more practical, it would be widely known and used by researchers and practitioners, even though researchers and practitioners do not always have the same perception of what "practical" is. Thus, while the PD approach introduced by Schwartz (1966) spawned several follow-up papers, to date, the classical inventory theory approach still predominates in the vast majority of the inventory management research literature and textbooks. The classical approach remains more popular, not only because of tradition, but also because it assigns a *direct* backorder/stockout cost, as opposed to its PD counterpart which assumes *indirect* costs. It is easier, quicker, and more familiar for a manager to think, "I want a customer service level of 90%," or equivalently "It costs me nine times more to allow backorders than to hold inventory" than to think in terms of the indirect stockout costs implied by the PD approach. How the manager picks the appropriate customer service level or the equivalent backorder cost, however, remains obscure. If operations management researchers are to continue teaching the classical inventory theory approach to students and advertising it to practitioners, however, they must continue seeking a credible answer to the question, what could the backorder penalty cost coefficient b be? At the same time, they must address the paradox of the classical approach mentioned above. In Chapter 2, we address these issues.

Schwartz's and other PD models take a macroscopic view at how the long-run average demand depends on customer service but fail to look at the dynamics of how this happens. There is a large body of operations management literature that studies the phenomenon whereby customers substitute one product with another or switch from one retailer to another when their first-choice product or retailer is out of stock. Yet, very few of these works look at the impact of such stockouts in future demand. In the short run, if a customer receives poor service from a supplier, he may defect to another supplier, and if he receives poor service from that supplier too, he may switch back to the original supplier. In other words, customers may switch from one supplier to another based on the service they receive from each supplier. Hence, the dynamic inventory control policy of one supplier may depend on the dynamic inventory control policy of the competing suppliers. This line of thought opens the path for game-theoretic formulations. In Chapter 3, we develop and analyze one such model.

Finally, although intuitively it makes sense that stockouts should affect future demand, no study on the impact of stockouts would be complete if it were not supported by empirical evidence. Recently, there has been an increasing call for rigorous empirical research in operations management. In contrast to other more mature management disciplines, operations management has the least developed empirical knowledge base to draw upon in answering challenging questions. This may be due to at least two reasons. Firstly, empirical research involves a systematic derivation and analysis of data from direct or indirect observation, a job that most operations management researchers are not well trained or interested in doing. Secondly, most companies that have the data are hesitant to share it with the rest of the world. In Chapter 4, we seek to find empirical evidence that stockouts do adversely affect future by performing a thorough statistical analysis of historical customer order and delivery data of a tool wholesaler and distributor over a period of four years.

1.2 Literature Review

The quantification of stockout costs that are related to the loss of customer goodwill has long been a difficult and unsatisfactorily resolved issue in the literature. As Gardner (1980) puts it, shortage cost parameters are no more real than the gods of Olympus. Nonetheless, the effects of stockouts on customer behavior have been studied quite extensively by the marketing research community. Most of the related work reported in the marketing research literature focuses on identifying and explaining consumer reaction to stockouts in retail settings. Such reaction may include item (brand and/or variety) or purchase quantity switching, cancellation or deferral of purchase, store switching, etc.

A number of studies postulate a decision model with alternative possible outcomes and courses of action of consumers and retailers following a stockout, and estimate the parameters (probabilities, costs, etc.) of the model via interviews and/or mail surveys.

Nielsen (1968a, b) documents the frequency of stockouts observed for items sold in supermarkets. In contrast to prior stockout studies that try to estimate the cost of a stockout on the basis of unsold inventory only, this study looks into consumer behavior. When recording stockouts, a distinction is made between availability of product on shelves and availability in the store, the latter meaning that the product is only available in the store backroom. The study also reports breakdowns for product categories, days of the week, levels of brand loyalty captured by certain product categories, and most importantly substitute, delay, or leave response.

Walter and Grabner (1975) design a model to describe the decision alternatives of a customer who encounters a stockout in a retail store, and conduct an empirical test of their model in liquor stores operated by the Ohio Department of Liquor Control.

Schary and Becker (1978) report the effects of a regional beer strike in which stockouts occurred in selected brands. Using brand share as the dependent variable, stockout effects are judged to be more short- than long-run. Schary and Christopher (1979) develop a model which identifies stockout response in relation to store and product decisions by consumers. They then compare this model to evidence of actual response to stockout situations collected at two units of a British supermarket chain. Their findings suggest that stockout perception is not universal and that reaction to stockouts influence the total image of the store.

Zinszer and Lesser (1981) look at how stockouts affect consumers of different demographic characteristics, whether the item was on sale and how the stockout affects store image and intended future patronage. Badinelli (1986) repeatedly ask decision makers to specify their marginal exchange rate between on-hand inventory and backorders, and then use the relatively more exact holding cost to estimate the shortage cost function through regression.

Emmelhainz et al. (1991) report the responses to an in-store interview of consumers who experience a stockout on items removed from the grocery shelves by researchers. They find that 32% of consumers purchase a different brand, 41% purchase a different size or variety of the same brand, and 14% go to another store.

Finally, Campo et al. (2000, 2004) investigate consumer reactions to stockouts – which are unexpected and temporary in nature – as opposed to permanent assortment reductions (PAR). Their results indicate that retailer losses incurred in case of a PAR may be substantially larger than those in case of a stockout for the same item. The results further suggest that stockout losses may disproportionately grow with OOS frequency and duration, emphasizing the need to keep their occurrence and length within limits.

Two exceptions of works that focus on business-to-business (B2B) rather than business-to-consumer (B2C) markets are Dion et al. (1991) and Dion and Banting (1995), who report the results of studies of the perceived consequences for B2B market buyers of being stocked out by their supplier and their repurchase loyalty on the next purchase occasion. The studies draw data from personal interviews and mail surveys. Buyers report lost sales and costly production disruptions resulting from the stockouts. The results show that buyers often seek an alternate supplier in the face of a stockout, but the majority return to the original supplier on the next purchase occasion.

Another group of marketing research studies is based on laboratory experiments.

Charlton and Ehrenberg (1976) is one example in which a panel of consumers in the UK is repeatedly offered the opportunity to buy certain artificial brands of a detergent. The study examines the effects of price differentials, a promotion, advertising, a stockout condition, the introduction of a new product, and certain weak forms of price differentiation on consumer dynamics, i.e. on how people change their purchasing habits. As far a the effects of the stockout condition is concerned, it is found that market shares and category sales return to their pre-stockout levels with no apparent long-term effects.

Motes and Castleberry (1985) repeat the same type of experiment using a real potato chip brand and find that market shares do not return to their pre-stockout levels whereas category sales do. Finally, Fitzsimons (2000) runs four laboratory experiments involving stockouts in a consumer choice context. The results of the experiments suggest that consumer response to stockouts is driven in large part by two factors: the effect of a stockout on the difficulty of making a choice from the set and the degree of personal commitment to the out-of-stock alternative.

There also exist a limited number of marketing studies that rely on historical data analysis.

Straughn (1991) is one of the first to use scanner data in a stockout study. She attempts to estimate the effects of stockouts on brand share for candy bars. The short-term effect is negligible. The long-term effect, defined as more than five weeks following the stockout condition, is substantial. Decline in brand share averages 10%.

Campo et al. (2003) explore the impact of retail stockouts on whether, how much and what to buy, by adjusting traditional purchase incidence, quantity and choice models, so as to account for stockout effects. Their study is based on scanner panel data of a large European supermarket chain.

There also exists a relatively recent survey- and experiment-based stream of research on consumers' perceptions of and reactions to waiting and service.

Anderson et al. (2006) conduct a large-scale field test with a national mail-order catalog and find that stockouts have an adverse impact on both the likelihood that a customer will place another order and the amount that the customer will spend on future orders (if any).

Taylor (1994) presents a model of the wait experience which assesses the effects of delay duration, attribution for the delay, and degree to which time is filled, on affective and evaluative reactions to the delay. An empirical test of the model with delayed airline passengers reveals that delays do affect service evaluations; however, this impact is mediated by negative affective reactions to the delay.

Carmon et al. (1995) examine how service should be divided and scheduled when it can be provided in multiple separate segments. They analyze variants of this problem using a model with a conventional function describing the waiting cost, which is modified to account for some aspects of the psychological cost of waiting in line. They show that consideration of the psychological cost can result in prescriptions that are inconsistent with the common wisdom of queuing theorists derived according to the conventional approach (e.g., equal load assignments).

Hui and Tse (1996) conduct an experimental study to examine the impact of two types of waiting information – waiting-duration information and queuing information – on consumers' reactions to waits of different lengths. Their results show that though acceptability of the wait and affective response to the wait have a significant mediating effect on the relationship between waiting information and service evaluation, perceived waiting duration does not. Moreover, neither type of information has significant impact in the short-wait condition, whereas waiting-duration information has greater impact than queuing information in the intermediate-wait condition and a smaller impact in the long-wait condition.

Kumar et al. (1997) examine the impact of the policy of a waiting time guarantee, on customers' waiting experiences and find that that a time guarantee, if met, increases satisfaction at the end of a wait; however, if violated, it decreases satisfaction at the end of the wait.

Zhou and Soman (2003) investigate consumers' affective experiences in a queue and their decisions to leave the queue after having spent some time in it (reneging). They find in their first two studies that, as the number of people behind increases, the consumer is in a relatively more positive affective state and the likelihood of reneging is lower.

Finally, Munichor and Rafaeli (2005) examine the effect of time perception and sense of progress in telephone queues on caller reactions to three tele-waiting time fillers: music, apologies, and information about location in the queue. In their first study, they find that call abandonment was lowest and call evaluations were most positive with information about location in the queue as the time-filler. In their second study, they find that the sense of progress in the queue rather than the perceived waiting time mediated the relationship between tele-waiting time filler and caller reactions. The issue of the value of time, and, to some extent, product availability seems to be a "hot" topic in marketing research today.

The effects of stockouts on customer behavior have also been studied by the operations management community. The first works that appeared in the operations management literature develop decision trees to model the consequences of stockouts.

Chang and Niland (1967) use decision trees to delineate the possible consequences of a stockout (e.g., lost sale, temporary cancellation of business, etc.) and their conditional probabilities of occurring, which they used to calculate an expected penalty cost. The main disadvantage of decision trees is the tremendous amount of time that is required for estimating their parameters. With this in mind, Oral et al. (1972) and Oral (1981) ask managers to define costs and probabilities for a stratified sample of their items. They then estimate the penalty costs for the remaining items by regressing the shortage cost on the gross unit profit.

Most of the research on the effects of stockouts on current and future sales in the operations management literature has focused on the development of mathematical inventory control models in which demand is presumed to be a function of a certain direct or indirect quantitative measure of stockouts, such as the average delivery delay, the percentage of demand that is satisfied from on-hand inventory, the percentage of immediately satisfied customers, etc.

Hanssmann (1959) is perhaps the first to model a relationship between inventory stocking policies and demand. He assumes that demand is normally distributed, with a constant coefficient of variation and a mean value that depends on the service level. He balances higher holding costs against increased sales in response to the decreased backlogging and the resulting reduction in delivery delays.

The concept of service level-dependent demand is further developed by Schwartz (1965, 1966, 1970), who develops an innovative “perturbed demand” model in which there is no fixed stockout cost but stockouts directly affect future demand. Our work in Chapter 2 relies heavily on Schwartz’s perturbed demand model.

Hill (1976) extends Schwartz's work by obtaining the optimal solution for a perturbed demand inventory model with fixed ordering and inventory holding costs. He restricts his attention to reorder quantity, reorder point policies, for which he concludes that the optimal reorder quantity is either identical to the EOQ and no stockouts are allowed, or is equal to the available storage capacity and the stockout level may be allowed to become rather large before an order is received.

Caine and Plaut (1976) come to the same conclusion as Hill (1976). Moreover, they obtain steady-state results for a stochastic periodic review model, similar to that studied by Schwartz (1970), where the demand is assumed to be follow an exponential distribution whose mean depends on the long run expected disappointment caused by stockouts rather than by the actual service received. They only look at a single-period problem with no cost on either ending inventory or backlogged demand.

Robinson (1991) provides a further generalization where the mean and variance of the demand in each period varies linearly with the number of satisfied customers in the previous period. He also gives an excellent literature review.

Argon et al. (2001) proposes a single item, periodic review model that investigated the effects of changes in the demand process that occur after stockout realizations. More specifically, they investigate a system where the demands in successive periods are deterministic but are affected by the backorder realizations.

Finally, there exist several works that study continuous-time, deterministic, inventory control systems in which the demand rate is assumed to be a polynomial function of the inventory level. A typical example of such a work is Urban (1995).

The above works remain within the framework of a single decision maker formulation and hence fail to look into the underlying competition interactions between suppliers. Given that the future (or even present) defection of a customer depends on what other options that customer has, several researchers address service-related issues within a game theoretic framework. There is a large body of operations literature works that study product and/or supplier substitution or switching when stockouts occur.

Li (1992) considers competition in production speed in a buyer's market, assuming that a demand will be filled by the supplier who produces the next available product first. This line of research has been followed and extended by other works. For example, Ernst and Cohen (1992) and Ernst and Powell (1995) consider a single-supplier, single-retailer system in which the demands faced by the retailer have a mean and standard deviation that depends on the steady state service level. Ernst and Powell (1998) model this system as a Stackelberg game with the supplier as the leader.

Lippman and McCardle (1997) introduces competition into the standard newsvendor problem. In their model, two firms make ordering decisions at the beginning of a period to compete for the demand in current period. When a shortage happens at one firm, the unmet demand switches to the other firm. Along the same line, Netessine et al. (2006) consider a two-firm competition problem in a reorder point system setting. When a stockout occurs at one firm, the unmet demand will either be backordered or switch to a competitor immediately. Stationary optimal ordering strategies are developed under four different scenarios. Since future demand is not affected by current activities, the problem is essentially a one-period problem.

Bernstein and Federgruen (2004a,b) consider price- and service-sensitive demands in a one-period setting, using a multiplicative demand model. They showed that the equilibrium in an infinite-horizon setting is the same as in the one-period setting.

Dana and Petruzzi (2001) consider a firm's inventory *and* price policy when it faces uncertain demand that depends on both inventory and price. They extend the classic newsvendor model by assuming that consumers who seek to maximize their expected utility chose between visiting the firm and consuming an exogenous outside option. The outside option represents the utility that the consumer forgoes when he chooses to visit the firm before knowing whether or not the product will be available. They investigate both cases of optimally chosen and exogenously given price.

From our literature review one can see that, since Schwartz (1966)'s work, the two factors, competition in product availability and its future effect have been more or less studied separately. The only exceptions are Gans (2002), Hall and Porteus (2000), Gaur and Park (2006), and Liu et al (2007).

More specifically, Gans (2002) models consumer learning and choice in response to random variation in the quality provided by competing suppliers. He develops an individual-level consumer demand model in which consumers use Bayesian updating to learn from their own experiences with the quality levels offered by suppliers. In each period, a consumer picks the supplier for which the consumer has the highest expectation of service level. Gans derives the steady-state characterization of this demand model when suppliers choose static quality policies and analyzes the competition between them.

Hall and Porteus (2000) investigate a simple dynamic model of firm behavior in which two firms compete by investing in capacity that is used to provide good service to their customers. There is a fixed total market of customers whose demands for service are random and who divide their patronage between the firms in each period. They address the issue of the firms' capacity decisions in response to customer service concerns.

Gaur and Park (2006) build on the model of Hall and Porteus (2000) by considering a model with asymmetric customer learning. When consumers experience positive or negative service encounters, they update their expectations about future encounters. Liu et al. (2007) also extend the work of Hall and Porteus (2000) by incorporating a more general demand model. In all three papers, however, it is assumed that there is no inventory carryover or backorder from period to period.

1.3 Thesis Organization

The remainder of this thesis is organized into three main parts which occupy Chapters 2-4, respectively.

Motivated by the need to determine reliable backorder penalty cost rates to be used in classical inventory control models, in Chapter 2, we revisit the classical Economic Order Quantity (EOQ) model with backorders that are being penalized with a backorder penalty cost coefficient, b, which reflects the intangible effect of the future loss of customer goodwill – and therefore demand – following a stockout. We ask the question, what could b be? To answer this question, we infer the value of b in the EOQ model with penalized backorders by connecting this model to Schwartz's (1966) seminal perturbed demand model which assumes that there is no explicit backorder penalty cost, but that the long-run average demand rate is an increasing function of the customer service level. The perturbed demand model proposed by Schwartz (1966) replaces the impracticable task of estimating b in the classical model with penalized backorders, with the more feasible task of estimating the parameters of the perturbed demand rate function.

In Chapter 3, we develop a newsvendor-type model of two suppliers that compete to sell the same type of items to a customer, repetitively, in discrete periods, for an infinite time horizon. At the beginning of each period, each supplier orders a number of items that are delivered to him immediately. In each period, the customer randomly chooses one of the two suppliers and demands from him a random number of items. The probability of choosing a supplier depends on the so-called "credibility level" of this supplier, which reflects the customer's estimate of the supplier's relative credibility based on the history of service – measured in terms of product availability – that both suppliers have provided to the customer in the past. The credibility levels of the suppliers change dynamically based on the quality of service – good or poor – that the customer receives in each period. We formulate the problem of finding optimal stationary ordering policies for both suppliers at equilibrium as a stochastic dynamic game, and we numerically solve the resulting optimality conditions for several instances of this problem. In all instances, the optimal ordering policy for each supplier turns out to be an order-up-to policy. Then, we restrict our attention to the case where each supplier has only two credibility levels, a low and a high, such that, when in the low level, he is never chosen by the customer, and when in the high level, he is always chosen by the customer. For this case, we assume that each supplier uses a credibility level-dependent order-up-to policy which leads to a Markov Decision Process with two decision makers. We numerically solve the resulting optimality conditions at equilibrium to find the optimal order-up-to levels of both suppliers.

Our work in Chapters 2 and 3 is based on developing and analyzing mathematical models in which it is assumed that stockouts affect future customer demand. In Chapter 4, we seek to find empirical evidence that such an effect exists. More specifically, we study the linkage between stockouts, customer service, current sales, and future demand, by performing a thorough statistical analysis of historical customer order and delivery data of a tool wholesaler and distributor over a period of four years. Our aim is find to evidence on the effect of stockouts on current sales and future customer demand in a wholesale business environment. We hope that the results of this analysis can provide useful information to operations management researchers who wish to develop and analyze realistic models of supplier-customer behavior, and at the same time complement our results in previous chapters. This analysis could also serve as an example for sales and inventory management practitioners who wish to perform a similar study on their own data. Another objective of Chapter 4 is to statistically analyze the customer order data itself. Given the lack of reports on real customer demand data in the literature, this analysis may be of particular interest to inventory management researchers who wish to develop and analyze realistic models of customer demand.

Finally, we summarize our findings in Chapter 5.

Chapter 2 Backorder Penalty Cost Coefficient "*b*": What Could It Be?

In this chapter, we revisit the Economic Order Quantity (EOQ) model with backorders that are being penalized with a backorder penalty cost coefficient, denoted by b, where b reflects the intangible effect of the future loss of customer goodwill following a stockout. We will henceforth refer to this model as the "penalized backorders" or "PB" model, for short. Our aim is to answer the question, what could b be in the PB model? To this end, we propose a scheme for inferring the value of b. This scheme is based on connecting the PB model to Schwartz's (1966) seminal alternative EOQ model with backorders and perturbed demand, which assumes that there is no explicit backorder penalty cost but that the long-run average demand rate is an increasing function of the fill rate. We will henceforth refer to the latter model as the "perturbed demand" or "PD" model, for short. The connection between the two models is accomplished by setting the optimal decision variables in the PB model equal to the respective variables in the PD model. While the idea of this connection and its implementation is the main contribution of this chapter, a secondary contribution is the exact analysis of the PD model that we provide along the way.

The rest of this chapter is organized as follows. In Section 2.1, we summarize some more or less known results on the optimal decision variables of the classical PB model and three variations of it in which we replace the explicit fixed order cost with a constraint on the order quantity, the interorder time, and the starting inventory in each cycle, respectively. We also perform a sensitivity analysis of the objective function of the classical PB model to our error in estimating b, in an attempt to answer the question, how crucial is it to know the true value of b? In Section 2.2, we derive analytical expressions for the optimal decision variables of the respective PD models, i.e., the classical PD model and three variations of it which correspond to the three variations of the PB model. In Section 2.3, we describe the scheme for inferring the backorder penalty cost coefficient b in the PB model from the PD model, and we use this scheme to infer b in the classical PB model and its three variations discussed in Section 2.1. We illustrate the scheme with a numerical example in Section 2.4, and we draw conclusions in Section 2.5.

2.1 The PB Model

In this section we revisit the classical PB model and three variations of it. The classical PB model assumes that a firm receives a single type of products from a supplier, holds them in inventory, and delivers them to its customers upon demand. The model also assumes that demand is continuous and constant over time, replenishment is instantaneous, delivery is immediate, the order quantity does not need to be an integer, and unfilled demand is backordered. Finally, the model assumes that the price margin per unit sold, the fixed order cost, the variable order cost per unit ordered, the inventory holding cost per unit stocked per unit time, and the backorder penalty cost per unit backordered per unit time are known and constant over time. The constant parameters of the PB model are denoted by the following symbols:

p = price margin (selling price minus purchase price) per unit sold;

k = fixed order cost;

h = inventory holding cost per unit stored per unit time;

b = backorder penalty cost per unit backordered per time unit;

D = demand per time unit.

The decision variables are the order quantity and the fraction of demand that is met from stock, or fill rate, which are denoted by the following symbols:

Q = order quantity;

F = fill rate.

The fill rate must satisfy $0 \leq F \leq 1$. Note that if $F = 0$, the firm operates in a pure make-to-order mode, backordering all the demand and not keeping any inventory. If $F = 1$, on the other hand, the firm operates in a pure make-to-stock mode, keeping inventory and not allowing any backorders. If $0 < F < 1$, the firm uses a mixed make-to-order and make-to-stock policy. The only constraint on the order quantity is that it must be nonnegative. In practical situations, however, there may be extra constraints on the decision variables which can lead to different variations of the PB model. In this chapter, we consider three such variations in which we replace the explicit fixed order cost with a constraint on the order quantity, the interorder time, and the starting inventory in each cycle, respectively. Figure 2-1 shows the inventory trajectory in the PB model.

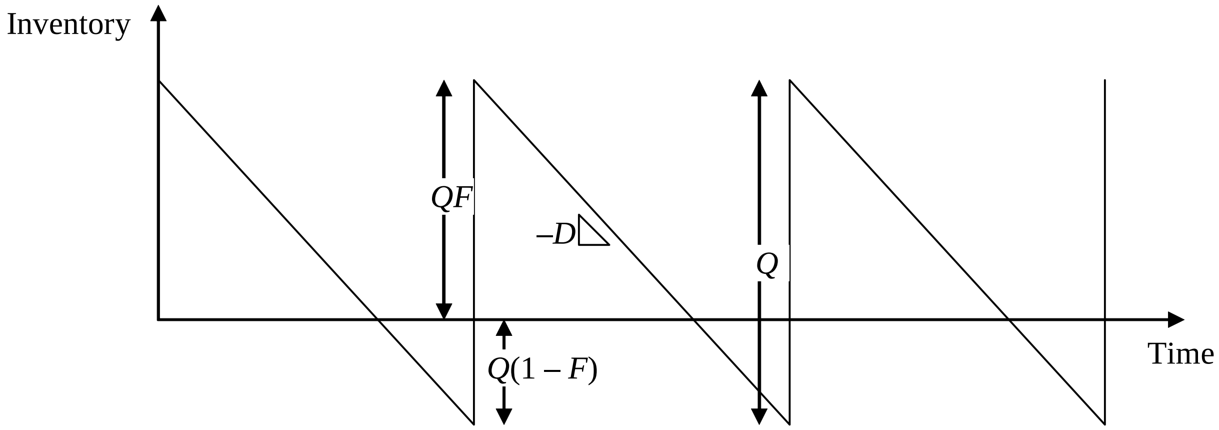

Figure 2-1: Inventory versus time for the PB model

The classical PB model and its three variations make up a total of four cases. For each case, it is straightforward to derive an expression of the average profit of the firm as a function of the decision variables Q and F. Table 2-1 shows the average profit function, denoted by the $P(Q, F)$, and the constraints for the four cases. Note that in all cases, the first term of the average profit represents the average reward, whereas the remaining terms represent the average cost. Also note that the decision variables Q and F only affect the average cost. The quantities Q, Q/D, and QF in the last column of Table 2-1 are the order quantity, the interorder time, and the starting inventory in each cycle, respectively, and $Q_{\min}$, $T_{\min}$, and $I_{\min}$ are positive, finite numbers denoting the minimum values of these quantities, respectively. Parameters $Q_{\min}$ and $T_{\min}$ may be set either externally by the supplier for economic reasons, or internally by the firm to incur an implicit fixed order expense, if the explicit fixed order cost k is not known or is difficult to obtain. Similarly, parameter $I_{\min}$ may be set internally by the firm to incur an implicit fixed order expense, or as a safety stock against fluctuations in demand, because in reality demand may vary.

Table 2-1 Objective function and constraints for the classical PB model and its three variations

#	$P(Q, F)$	Constraints
1	$pD - k\dfrac{D}{Q} - h\dfrac{QF^2}{2} - b\dfrac{Q(1-F)^2}{2}$	$0 \le F \le 1$, $Q \ge 0$
2	$pD - h\dfrac{QF^2}{2} - b\dfrac{Q(1-F)^2}{2}$	$0 \le F \le 1$, $Q \ge Q_{\min}$
3	$pD - h\dfrac{QF^2}{2} - b\dfrac{Q(1-F)^2}{2}$	$0 \le F \le 1$, $Q/D \ge T_{\min}$
4	$pD - h\dfrac{QF^2}{2} - b\dfrac{Q(1-F)^2}{2}$	$0 \le F \le 1$, $QF \ge I_{\min}$

In all four cases of the PB model, the objective is to find the optimal values of the decision variables Q and F that maximize the average profit subject to the constraints. The methodology to do this is quite standard and consists of the following four steps: (1) Express the optimal order quantity as a function of F, say $Q^*(F)$, (2) write an expression for the average profit as a function of F only, say $P^*(F)$, after having replaced Q by $Q^*(F)$, i.e., $P^*(F) = P(Q^*(F), F)$, (3) maximize $P^*(F)$, subject to $0 \le F \le 1$, to determine the optimal fill rate F^*, and (4) evaluate $Q^*(F^*)$ to determine the optimal order quantity Q^*. The implementation of these steps can be found in many textbooks on inventory management (e.g. Zipkin, 2000), at least for the classical PB model (case 1). For cases 2-4, it can be easily carried out in a similar manner, so we omit the details here for space considerations. The results for all the cases are summarized in Table 2-2, where P^* in the last column is the optimal average profit, i.e., $P^* = P^*(F^*) = P(Q^*(F^*), F^*) = P(Q^*, F^*)$.

Table 2-2: Optimal decision variables and objective function for the classical PB model and its three variations

#	$Q^*(F)$	$P^*(F)$	F^*	Q^*	P^*
1	$\sqrt{\dfrac{2kD}{hF^2+b(1-F)^2}}$	$pD-\sqrt{2kD\left[hF^2+b\left(1-F\right)^2\right]}$	$\dfrac{b}{h+b}$	$\sqrt{\dfrac{2kD(h+b)}{hb}}$	$pD-\sqrt{\dfrac{2kDhb}{(h+b)}}$
2	$Q_{\min}$	$pD-h\dfrac{Q_{\min}F^2}{2}-b\dfrac{Q_{\min}(1-F)^2}{2}$	$\dfrac{b}{h+b}$	$Q_{\min}$	$pD-\dfrac{hbQ_{\min}}{2(h+b)}$
3	$DT_{\min}$	$pD-h\dfrac{DT_{\min}F^2}{2}-b\dfrac{DT_{\min}(1-F)^2}{2}$	$\dfrac{b}{h+b}$	$DT_{\min}$	$pD-\dfrac{hbDT_{\min}}{2(h+b)}$
4	$\dfrac{I_{\min}}{F}$	$pD-h\dfrac{I_{\min}F}{2}-b\dfrac{I_{\min}(1-F)^2}{2F}$	$\sqrt{\dfrac{b}{h+b}}$	$I_{\min}\sqrt{\dfrac{h+b}{b}}$	$pD-\left(\sqrt{b\left(h+b\right)}-b\right)I_{\min}$

From column 4 of Table 2-2, we can observe that in all four cases, the optimal fill rate F^* is a function of the backorder penalty cost coefficient b. More specifically, in cases 1-3, F^* is given by the familiar fraction $b/(h + b)$, whereas in case 4, it is given by the square route of this fraction. From Tables 2-1 and 2-2, it is easy to see that if $Q_{\min} = DT_{\min}$, cases 2 and 3 are identical to each other. This means that there are really only three cases of the PB model to consider; however, we purposely leave the results for both cases 2 and 3 in Table 2-2, because in Section 2.3 we will relate them to the results of the respective cases of the PD model, which, as we will see then, are not identical to each other. From Table 2-2, it is also easy to see that in cases 2-4, the average profit $P(Q, F)$ is strictly decreasing in the order quantity Q and that Q is only restricted by a lower limit.

For this reason, the optimal order quantity Q^* is simply set at this lower limit, as can be seen in column 5 of Table 2-2. In cases 2 and 3, this limit is independent of F and is equal to $Q_{\min}$ and $DT_{\min}$, respectively. In case 4, it is given by $I_{\min}/F$, which becomes $I_{\min}/F^*$ once the optimal fill rate F^* is set. Finally, case 1 is the only case where $P(Q, F)$ is not strictly decreasing in Q, because of the extra fixed order cost term, $-kD/Q$, which is increasing in Q. In this case, the optimal order quantity is given by the familiar square root formula in Table 2-2.

To summarize, in all cases, F^* is a function of b. Moreover, in cases 1 and 4, Q^* is a function of F^*, and hence also a function of b. In cases 2 and 3, on the other hand, Q^* does not depend on b.

Given the difficulty in estimating b, a natural question that first comes to mind is how crucial is it to know the true value of b? To answer this question, we will investigate the sensitivity of the average cost to our *error* in estimating b for the classical PB model (case 1). Note that this investigation is different from examining the sensitivity of the average cost to a *change* in b, which is carried out in Zipkin (2000). To carry out our investigation, suppose that instead of the true backorder penalty cost coefficient b, we use a wrong backorder penalty cost coefficient, say b'. It is straightforward to show that the ratio of the optimized average cost using the wrong coefficient b' over the true optimal average cost using the true coefficient b is equal to

$$\sqrt{\frac{(1+\alpha)}{\beta(1+\alpha\beta)}}\frac{1+\beta+2\alpha\beta^2}{2(1+\alpha\beta)} \tag{2.1}$$

where α is defined as the ratio of the true backorder penalty cost coefficient over the inventory holding cost, i.e., $\alpha = b/h$, and β is defined as the ratio of the wrong backorder penalty cost coefficient over the true backorder penalty cost coefficient, i.e., $\beta = b'/b$. It can be easily verified that when $\beta = 1$, the cost ratio given by expression (2.1) is equal to 1, whereas when $\beta \neq 1$, it is strictly greater than 1, as is normally expected. A noteworthy observation is that the ratio of the two costs given by (2.1) is a function of the unitless coefficients α and β only and does not depend on the absolute values of h and b, the demand rate D, or the fixed order cost k. In other words, the relative cost of making an error in estimating b depends only on the size of this error, which is given by β, and on the ratio of the true backorder penalty cost coefficient over the inventory holding cost, which is given by α. To gain more insight on expression (2.1) we evaluated it numerically for different values of α and β. The results are shown in Table 2-3.

The first row of Table 2-3 shows that if $\beta = 0.1$, i.e., if we use a wrong backorder penalty cost coefficient b' which is ten times smaller than the true coefficient b, the optimized average cost using b' is bigger than the optimal average cost (using b) by a factor which ranges from 1.8004, when $\alpha = 0.1$, to 2.4103, when $\alpha = 10$. Similarly, the last row of Table 2-3 shows that if $\beta = 10$, i.e., if we use a wrong coefficient b' which is ten times bigger than the true coefficient b, the optimized average cost using b' is bigger than the optimal average cost (using b) by a factor which ranges from 1.8175, when $\alpha = 0.1$, to 1.0390, when $\alpha = 10$. These observations suggest that the average cost is not very sensitive to errors in estimating the backorder penalty cost coefficient. This is not too surprising, because using the wrong backorder penalty cost coefficient essentially leads to the computation of suboptimal values of the decision variables Q and F, and it is a well-known result in Inventory Theory that the objective function of the PB model is not very sensitive to the decision variables Q and F.

Table 2-3: Ratio of the optimized average cost using b' over the true optimal average cost using b, for different values of α and β

	α				
β	0,1	0,5	1	2	10
0,1	1,80	2,00	2,17	2,38	2,41
0,5	1,07	1,08	1,09	1,08	1,04
1	1,00	1,00	1,00	1,00	1,00
2	1,07	1,07	1,06	1,04	1,01
10	1,82	1,46	1,29	1,17	1,04

A less obvious observation from the results of Table 2-3 is that the cost ratio increases with α, if $\beta = 0.1$, whereas it decreases with α, if $\beta = 10$. This suggests that underestimating b, if b is already much smaller than h, is not as bad as underestimating b, if b is much larger than h. Similarly, overestimating b, if b is already much larger than h, is not as bad as overestimating b, if b is much smaller than h. This distinction is important to note in light of the fact that in most practical situations, backorders are deemed by managers to be much more expensive than inventories, i.e., b is judged to be much larger than h. An exception of this is in industries where there is rapid spoilage or obsolescence of inventories.

2.2 The PD Model

The sensitivity analysis in the previous section showed that in the classical PB model, the average cost is fairly robust to errors in estimating b. This is somewhat of a relief, because it means that not picking the right value for *b* is not so catastrophic; however, it still does not answer the question, what could *b* be?

The difficulty in estimating *b* lies in the fact that *b* is supposed to reflect the intangible effect of the loss of customer goodwill following a stockout. As Schwartz (1966) noted, however, the effect of the loss of goodwill should not be a penalty cost of the type considered in the PB model. This is because the backorder penalty cost term in the objective function in all the cases of the PB model in Table 2-1 is subtracted as though the firm incurs an expense at the time of the stockout. Yet, the effect of the loss of customer goodwill is incurred not at the time of the stockout incident, but at a later time, due to the customer's decision to alter his future demand. With this in mind, Schwartz (1966) modified the PB model by eliminating the explicit backorder penalty cost term from the objective function and by assuming that the long-run demand rate – and hence the long-run average reward – is an increasing function of the fill rate. Schwartz called the resulting model a "perturbed demand" model. As was mentioned in Section 2.1, in this chapter, we will refer to Schwartz's model as the "PD" model. More specifically, Schwartz proposed the following long-run demand rate function in his PD model:

$$D'(F') = \frac{A}{1+\left(1-F'\right)B} \tag{2.2}$$

where F' is the long-run average fill rate, and parameters A and B are defined as follows:

A = maximum potential demand rate corresponding to a fill rate equal to one;

B = reflects the long-run average amount that the customer does not buy because of his disappointment due to stockouts.

Note that we used the symbols "F'" and "$D'(\cdot)$" to represent the fill rate and the demand rate in the PD model, in order to distinguish these variables from the respective variables in the PB model. In fact, in this chapter, as a rule, we will use the symbol "prime" for those variables and functions in the PD model that also appear in the PB model, to make the distinction between the two models clear. Note that the perturbed demand rate $D'(F')$ given by (2.2) is constant, since it refers to a constant long-run, average fill rate F'.

To better see how parameter B affects $D'(F')$, we evaluated the ratio of the demand rate when the fill rate is F' over the maximum demand A, versus F', for different values of B, and we plotted the results in Figure 2-2. From Figure 2-2, we can see that when B = 0, the demand rate remains constant as the fill rate F' drops from 1 to 0. When B = 1, the demand rate drops to half of its maximum value as F' drops from 1 to 0. When $B = 9$, the demand rate drops to only 1/10 of its maximum value as F' drops from 1 to 0.

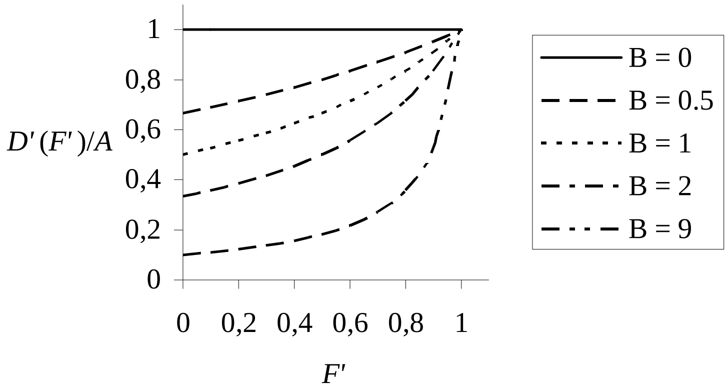

Figure 2-2: $D'(F')/A$ vs. F' for different values of B

The PD model proposed by Schwartz (1966) replaces the impracticable task of estimating b in the PB model, with the more feasible task of estimating parameters A and B of the perturbed demand rate function $D'(F')$ given by (2.2). Schwartz (1966) also proposed a procedure for measuring parameters A and B from observed demand data.

In a follow-up paper, Schwartz (1970) continued his investigation of the PD model by formulating three variations of this model that correspond to the three variations of the PB model discussed in the previous section. In fact, he also considered the same variations with lost sales instead of order backlogging, but we will not discuss these variations in this chapter for space considerations. If to the three variations with backlogging we add another variation which corresponds to the classical PB model and call this the "classical" PD model, then we have a total of four cases of the PD model which correspond to the four cases of the PB model discussed in Section 2.1. Table 2-4 shows the average profit function, denoted by the $P'(Q', F')$, and the constraints for these four cases, where Q' and F' are the decision variables.

In the three variations of the PD model that Schwartz (1970) considered, i.e., cases 2-4 of Table 2-4, he merely stated in 7-8 lines the first-order condition for the optimal quantity of unfilled demand which he denoted by L, i.e., in Schwartz's model $L = Q'(1 - F')$. However, in none of these cases, except case 2, did he solve this condition or provide any further analysis, discussion, or insights. In other words, Schwartz formulated the models for cases 3 and 4 but did not solve them. In case 2, Schwartz stated the first-order condition – a cubic equation – for the optimal value of L and then followed it with some observations obtained from a "simple graphical analysis" to describe the roots of the optimality condition. However, the "simple graphical analysis" that Schwartz invoked to describe the roots of the optimality condition is hardly a proof and in fact does not cover the entire range of the model parameters. Also, Schwartz's subsequent proposal to "solve the cubic [condition] numerically" to determine the optimal L does not provide any analytical results and insights about the solution.

Table 2-4: Objective function and constraints for the classical PD model and its three variations

#	$P'(Q', F')$	Constraints
1	$pD'(F') - k\dfrac{D'(F')}{Q'} - h\dfrac{Q'F'^2}{2}$	$0 \le F' \le 1$, $Q' \ge 0$
2	$pD'(F') - h\dfrac{Q'F'^2}{2}$	$0 \le F' \le 1$, $Q' \ge Q_{\min}$
3	$pD'(F') - h\dfrac{Q'F'^2}{2}$	$0 \le F' \le 1$, $Q'/D'(F') \ge T_{\min}$
4	$pD'(F') - h\dfrac{Q'F'^2}{2}$	$0 \le F' \le 1$, $Q'F' \ge I_{\min}$

In this chapter, we analytically solve the underlying optimization problems corresponding to the four cases of the PD model, and for each case we produce precise expressions and conditions for the optimal decision variables. The methodology to do this is quite standard and consists of the same four steps for solving the four cases of the PB model, outlined in Section 2.1. However, implementing this methodology on the PD model is much more complicated than implementing it on the PB model, because the perturbed demand rate function $D'(F')$ given by (2.2) complicates the average profit function $P'(Q', F')$ shown in column 2 of Table 2-4. The details of this implementation can be found in the Appendix of this chapter. The results of the implementation of the first two steps of the methodology are summarized in Table 2-5, whereas the results of the implementation of the last two steps are summarized in Table 2-6, where F_2' is the smallest real root of

$$\frac{pAB}{\left[1+(1-F')B\right]^2} - hQ_{\min}F' \tag{2.3}$$

and

$$F_3' = 1 + \frac{1}{B} - \sqrt{\frac{(1+B)^2}{B^2} - \frac{2p}{h^2T_{\min}^2}} \tag{2.4}$$

Table 2-5: Optimal order quantity and average profit as a function of F' for the classical PD model and its three variations

#	$Q'^{*}(F')$	$P'^{*}(F')$
1	$\sqrt{\dfrac{2kA}{hF'^2\left[1+(1-F')B\right]}}$	$\dfrac{pA}{1+(1-F')B}-F'\sqrt{\dfrac{2khA}{1+(1-F')B}}$
2	$Q_{\min}$	$\dfrac{pA}{1+\left(1-F'\right)B}-h\dfrac{Q_{\min}F'^2}{2}$
3	$\dfrac{AT_{\min}}{1+\left(1-F'\right)B}$	$\dfrac{pA}{1+\left(1-F'\right)B}-\dfrac{hAT_{\min}F'^2}{2\left[1+\left(1-F'\right)B\right]}.$
4	$\dfrac{I_{\min}}{F'}$	$\dfrac{pA}{1+\left(1-F'\right)B}-h\dfrac{I_{\min}F'}{2}$

Table 2-6: Optimal decision variables for the classical PD model and its three variations and inferred backorder penalty cost coefficient and resulting optimal order quantity for the respective PB models

	PD Model			PB Model	
#	F'^{*}	Q'^{*}	Condition	b	Q^{*}
1	0	∞	$p\sqrt{A}B/(1+B)<\sqrt{2hk}$	0	∞
	1	$\sqrt{\dfrac{2kA}{h}}$	$p\sqrt{A}B/(1+B)>\sqrt{2hk}$	∞	$\sqrt{\dfrac{2kD}{h}}$
	$0, 1$	$\infty, \sqrt{\dfrac{2kA}{h}}$	$p\sqrt{A}B/(1+B)=\sqrt{2hk}$	$0, \infty$	$\infty, \sqrt{\dfrac{2kD}{h}}$
2	F_2'	$Q_{\min}$	$pAB<hQ_{\min}$ or $pAB=hQ_{\min}, B>0.5$ or $pAB>hQ_{\min}, B>0.5,\ P'^{*}(F_2')>pA-hQ_{\min}/2$	$h\dfrac{F_2'}{1-F_2'}$	$Q_{\min}$
	1		$pAB\geq hQ_{\min}, B\leq 0.5$ or $pAB>hQ_{\min}, B>0.5, P'^{*}(F_2')<pA-hQ_{\min}/2$	∞	
	F_2', 1		$pAB>hQ_{\min}, B>0.5, P'^{*}(F_2')=pA-hQ_{\min}/2$	$h\dfrac{F_2'}{1-F_2'}$ ∞	
3	F_3'	$\dfrac{AT_{\min}}{1+\left(1-F_3'\right)B}$	$2pB/(2+B)<hT_{\min}$	$h\dfrac{F_3'}{1-F_3'}$	$DT_{\min}$
	1	$AT_{\min}$	$2pB/(2+B)\geq hT_{\min}$	∞	
4	0	∞	$2pAB/(1+B)<hI_{\min}$	0	∞
	1	$I_{\min}$	$2pAB/(1+B)>hI_{\min}$	∞	$I_{\min}$
	$0, 1$	$\infty, I_{\min}$	$2pAB/(1+B)=hI_{\min}$	$0, \infty$	$\infty, I_{\min}$

Columns 2-4 of Table 2-6 show the optimal decision variables and the conditions under which they hold, for the four cases of the PD model. The last two columns of Table 2-6 show the inferred backorder penalty cost coefficient b and the resulting optimal order quantity Q^* in the respective cases of the PB model. These two columns were inserted in Table 2-6 for space considerations, but they will be discussed in Section 2.3, where they belong.

From the results shown in columns 2-4 of Table 2-6, we can observe that cases 1 and 4 of the PD model are somewhat similar to each other, and so are cases 2 and 3. This was also true for the respective cases of the PB model, discussed in Section 2.1.

The most striking similarity between cases 1 and 4 of the PD model is that in both cases the optimal fill rate F'^* is always either zero or one. This means that in these two cases, it is always optimal either to only hold inventory and not allow any backorders ($F'^*= 1$), thus maximizing the long-run demand, or to only allow backorders and not hold any inventory ($F'^*= 0$), thus minimizing the long-run demand. Holding inventory and allowing backorders is never optimal. This result was not obvious a priori. In fact, in the respective PB models discussed in Section 2.1, we would have hardly thought that such a result might hold. Still, this result is not irrational. We know from experience that there are some companies that do not tolerate any backorders, while other companies operate in a pure make-to-order mode. The reason that F'^* is always either zero or one in cases 1 and 4 of the PD model is that the average profit function $P'^*(F')$ for these two cases, shown in the last column of Table 2-5, is always maximized at one or both of the end points of the interval of permissible values of F', [0, 1], as is shown in the Appendix of this chapter.

On the other hand, in both cases 2 and 3 the optimal fill rate F'^* is always either one or equal to a quantity which is less than one but strictly greater than zero, even if the inventory holding cost rate h is extremely large, as long as it is not infinite. This quantity is denoted by F_2' and F_3', for cases 2 and 3, respectively, and depends on the model parameters; therefore, it can assume a continuum of values, depending on these parameters. This means that in these two cases, it is always optimal either to only hold inventory and not allow any backorders ($F'^*= 1$), or to allow backorders for some time and hold inventory for the rest of the time ($0 < F'^* < 1$). The reason that F'^* can never be zero in cases 2 and 3 of the PD model is that in these two cases, the first derivative of the average profit $P'^*(F')$ shown in Table 2-5 is always positive at $F' = 0$, as is shown in the Appendix of this chapter. This implies that it is always preferable to keep some inventory – as low as it may be – than not to keep any inventory at all.

The only interpretation that we can give about why in cases 1 and 4, F'^* is zero whereas in cases 2 and 3 it assumes a continuum of finite values is the following. In cases 1 and 4, F'^* *can* be zero, because Q'^* *can* go to infinity, and when Q'^* goes to infinity, F'^* *must* be zero, as we will explain latter. In cases 2 and 3, on the other hand, Q'^* *can not* go to infinity, and therefore F'^* *does not have* to be zero (in fact, as we mentioned in the previous paragraph, it *can not* be zero, because the first derivative of the average profit $P'^*(F')$ is always positive at $F' = 0$) but instead assumes a continuum of finite values. We should note, however, that if, in cases 1 and 4, we impose a finite upper limit, say $Q_{\max}$, on the order quantity, then it can be shown that F'^* *can not* be zero but instead assumes a continuum of positive values, just like in cases 2 and 3. At the same time, if the minimum order quantity $Q_{\min}$ in case 2, or the minimum interorder time $T_{\min}$ in case 3, tends to infinity, then Q'^* will also tend to infinity, as we will explain latter, and therefore F'^* will tend to zero, just like in cases 1 and 4.

For the cases where F'^* is always zero or one, i.e., cases 1 and 4, the decisive condition of whether to only hold inventory ($F'^* = 1$) or only allow backorders ($F'^* = 0$) is $p\sqrt{A}B/(1+B) < \sqrt{2hk}$, in case 1, and $2pAB/(1 + B) < hI_{\min}$, in case 4. From these conditions, we can see that in both cases, increasing the reward related parameters p, A, or B, tends to favor the solution $F'^* = 1$, i.e., only hold inventory. On the other hand, increasing the cost related parameters h or k, in case 1, and similarly increasing h or $I_{\min}$, in case 4, tends to favor the solution $F'^* = 0$, i.e., only allow backorders.

In case 1, the parameter that affects mostly the decisive condition is the price margin p, because it appears linearly in this condition, whereas parameters h, k and A are in a square root, and parameter B appears in a term that ranges between zero and one. In case 4, on the other hand, parameters p, A, h, and $I_{\min}$ affect equally strongly the decisive condition, because they appear linearly in this condition. In contrast, the effect of parameter B is weaker, because B appears in a term that ranges between zero and one. To better understand this point, think of an instance where the left-hand side (lhs) of the decisive condition of case 1 is slightly greater than half of the right-hand side (rhs). Then, from the results in Table 2-6, $F'^* = 0$, i.e., the firm should operate in a pure make-to-order mode. Suppose that the firm can double the price margin p without affecting the demand rate. Then, the lhs of the decisive condition will become slightly larger than the rhs. Again, from the results in Table 2-6, this means that $F'^* = 1$, i.e., the firm should switch its operation from a pure make-to-order to a pure make-to-stock mode. If the firm did not have the option of changing the price margin p, however, then in order to achieve the switch from make-to-order to make-to-stock, which could be achieved by doubling p, it would have to quadruple the maximum demand rate A. The reason for which in case 1 the decisive condition is more sensitive to p than to the other parameters is that p appears linearly in the objective function $P'^*(F')$, whereas the other parameters appear in a square root (see column 3 of Table 2-5).

From columns 2 and 3 in Table 2-6, we can observe that in all cases, if $F'^* = 1$, then Q'^* is finite. If $F'^* = 0$, however, which is true only in cases 1 and 4, then $Q'^* = \infty$. The reason for this is slightly different in each case. More specifically, in both cases, the appropriate decisive condition determines whether $F'^* = 1$ or $F'^* = 0$. The tradeoff at stake, favoring one or the other solution, is between incurring high inventory costs (and, in case 1, high ordering costs as well) on one hand, and losing long-term demand and therefore revenue, on the other hand. If the model parameters in the decisive condition dictate the solution $F'^* = 0$, then it is optimal for the firm to operate strictly with backorders and no inventory. Since backorders incur no direct cost, the firm can have as many of them as it pleases for free. This much is true for both cases 1 and 4. The difference about why $Q'^* = \infty$, between the two cases, is that in case 1, given that every time the firm orders a quantity Q', it pays an order cost k, then why not have Q' be infinite to avoid paying the order cost? Hence, $Q'^* = \infty$. In case 4, on the other hand, if $F'^* = 0$, then Q'^* must be infinite, not to avoid paying the order cost, since there is not any, but because otherwise, the minimum-inventory constraint $Q'F' \geq I_{\min}$ will be violated. Of course, in real life, the order quantity can not be infinite. This can be handled in the model by assuming that the order quantity has an upper limit, say $Q_{\max}$, which is large enough so that $Q_{\max} \geq \sqrt{2kA/h}$, in case 1, and $Q_{\max} \geq I_{\min}$, in case 4, and then resolving the optimization problem with the additional constraint $Q' \leq Q_{\max}$ to obtain F'^*. As was mentioned earlier, it can be shown that if we impose such a limit, F'^* can not be zero but instead assumes a continuum of positive values, just like in cases 2 and 3.

For the cases where F'^* is always either one or between zero and one, i.e., cases 2 and 3, the decisive condition of whether to only hold inventory ($F'^* = 1$) or allow backorders for some time and hold inventory for the rest of the time ($0 < F'^* < 0$) is as follows. In case 2, the decisive condition is complicated and can be analyzed into three levels of subconditions. These subconditions are $pAB < hQ_{\min}$, at the first level, $B > 0.5$, at the second level, and $P'^*(F_2') > pA - hQ_{\min}/2$, at the third level. More specifically, if $pAB < hQ_{\min}$, then the inventory holding cost is high enough compared to the loss of revenue caused by a drop in long-term demand, so that the firm can afford to allow some backorders, no matter how small B is, as long as it is not zero; hence $F'^* < 1$. If

$pAB \geq hQ_{\min}$, however, then the inventory holding cost may not be high enough compared to the loss of revenue caused by a drop in long-term demand to always allow some backorders. In this case, whether to allow some backorders or not depends on the value of B. Namely, if $B \leq 0.5$, the revenue term in $P'^{*}(F')$ always increases faster than the inventory cost term, as F' increases from 0 to 1, and therefore, $F'^{*} = 1$, i.e., no backorders are allowed (see Appendix of this chapter). If $B > 0.5$, on the other hand, the revenue term in $P'^{*}(F')$ increases faster than the inventory cost term, as F' increases from 0 to the smallest root of the derivative of $P'^{*}(F')$, F'_2, then the reverse is true as F' increases from F'_2 to the second smallest root, and finally the revenue term increases faster than the inventory cost term again, as F' increases from the second smallest root to 1. In this case, the optimal fill rate depends on whether $P'^{*}(F'_2)$ or $P'^{*}(1)$ is larger.

From the first subcondition of case 2, $pAB < hQ_{\min}$, we can see that increasing parameters p, A, or B, tends to favor the solution $F'^{*} = 1$, i.e., hold inventory and do not allow any backorders, whereas increasing h or $Q_{\min}$ tends to favor the solution $F'^{*} = F'_2 < 1$, i.e., hold inventory but also allow some backorders. We can also see that all five parameters affect the first decisive condition equally strongly, because they all appear linearly in this condition. Finally, increasing the minimum order quantity $Q_{\min}$, decreases the smallest real root of expression (2.3) and hence F'_2. In fact, as $Q_{\min}$ tends to infinity, F'_2 tends to zero. This is because, as $Q_{\min}$ tends to infinity, the firm is obliged to order a quantity that tends to infinity. If it keeps this quantity in stock, its inventory holding cost will also tend to infinity. To avoid this, it is preferable for the firm to backorder this quantity and pay the price of a reduced long-run demand rate, which is certainly finite.

In case 3, the decisive condition of whether to only hold inventory or allow backorders for some time and hold inventory for the rest of the time is $2pB/(2 + B) < hT_{\min}$. From this condition, we can see that, similarly to case 2, increasing parameters p or B, tends to favor the solution $F'^{*} = 1$, whereas increasing h or $T_{\min}$ tends to favor the solution $F'^{*} = F'_3 < 1$. Moreover, parameters p, h, and $T_{\min}$ affect equally strongly the decisive condition, because they appear linearly in this condition, whereas the effect of parameter B is weaker, because B appears in a term that ranges between zero and one. Unlike, case 2, and for this matter cases 1 and 4 as well, in case 3, the maximum potential demand rate A is missing from the decisive condition as well as from the expression for F'_3 given by (2.4). This is because A appears linearly in all the terms of the average profit $P'^{*}(F')$ shown in column 3 of Table 2-5 and in the Appendix of this chapter; therefore, all A does is simply multiply $P'^{*}(F')$ and its derivative by a constant without really affecting their roots.

Finally, it can be seen from (2.4) that increasing the minimum interorder time $T_{\min}$, decreases F_3'. In fact, as $T_{\min}$ tends to infinity, F_3' tends to zero, essentially for the same reason that F_2' tends to zero, as $Q_{\min}$ tends to infinity, explained in the preceding paragraph.

Finally, recall from our discussion in Section 2.1, that cases 2 and 3 of the PB model are identical to each other, if $Q_{\min} = DT_{\min}$. This is no longer true for cases 2 and 3 of the PD model, because the demand rate is not a constant, as was the case in the PB model, but a function of the fill rate F'.

2.3 Inferring *b* in the PB Model from the PD Model

In the last sentence of his conclusions, Schwartz (1970) wrote, “The Perturbed Demand approach to goodwill stockout penalties is both substantially more valid and more practical than any previously considered in the literature of inventory theory.” We agree with Schwartz that the PD approach to goodwill stockout penalties is more valid than the classical inventory theory approach, for two reasons. The first reason is the already stated difficulty in picking a good – let alone the best – value for the backorder (or stockout) coefficient or the customer service level in the classical approach. The second reason is that the classical approach has the following paradox embedded in it. It supposes that there is a backorder penalty cost which reflects the future loss of demand due to the loss of customer goodwill following stockouts, and yet it assumes that the demand is stationary.

While we find that the PD approach to goodwill stockout penalties is more valid than the classical inventory theory approach, we are not sure if it is more practical than the classical approach. If it were more practical, it would be widely known and used by researchers and practitioners, even though researchers and practitioners do not always have the same perception of what “practical” is. Thus, while the PD approach introduced by Schwartz (1966, 1970) spawned several follow-up papers, to date, the classical inventory theory approach still predominates in the vast majority of the inventory management research literature and textbooks. The classical approach remains more popular, not only because of tradition, but also because it is more convenient to use by managers, as it assigns a *direct* backorder/stockout cost, instead of the *indirect* costs implied by its PD counterpart. It is easier, quicker, and more familiar for a manager to think, “I want a customer service level of 90%,” or equivalently “It costs me nine times more to allow backorders than to hold inventory” than to think in terms of the indirect stockout costs implied by the PD approach. In addition, most ERP systems and other decision support tools used in practice rely on the input of user-defined customer service levels. How a manager picks the appropriate customer service level or the equivalent backorder cost for an item, however, remains obscure. If OM researchers are to continue teaching the classical inventory theory approach to students and advertising it to practitioners, they must continue seeking a credible answer to the question, what could the backorder penalty cost coefficient b be? At the same time, they must address the paradox of the classical approach mentioned above.

With this in mind, in this section, we propose a scheme for inferring the value of b in the PB model which is based on connecting the PB model to Schwarz's PD model. The connection between the two models is accomplished by asking the question, what should b be in the PB model to make the optimal decision variables Q^* and F^* in this model identical to the optimal decision variables Q'^* and F'^* in the PD model, for given parameters A and B and the same reward and cost parameters p, k, and h, as those used in the PB model?

The sought after *inferred value* of b in the PB model should not be regarded as an instantaneous, explicit expense that the firm incurs at the time of a stockout. Rather, it should be thought of as an artificial, implicit backorder penalty cost coefficient that the firm should use in order to maximize its long-run average profit. To elaborate more on this line of thought, the average demand rate D in the PB model should not be regarded as a constant, long-run average demand rate, but as an average demand rate which is constant only in the short run. This is because as time passes, no matter what the initial value of D is, the average demand rate will drift towards $D'(F)$, assuming that the fill rate F is kept constant, so that in the long run, its average value will be equal to $D'(F)$. This further means that the average profit in the PB model given in column 2 of Table 2-1 should be regarded as a short-run rather than as a long-run average profit. By using the optimal decision variables Q^* and F^* in the PB model, the firm is therefore maximizing its "artificial" short-run average profit which is given in column 2 of Table 2-1. At the same time, however, it is also shaping its true long-run average profit which is given in column 2 of Table 2-4, because the fill rate F^* that it is using determines the long-run average demand rate through equation (2.2). We used the adjective "artificial" to describe the average profit given in column 2 of Table 2-1 in the PB model, because this profit depends on b, which as we mentioned earlier, should be thought as an artificially-set parameter. To complete our reasoning, if the optimal decision variables Q^* and F^* in the PB model are identical to the optimal decision variables Q'^* and F'^* in the PD model, then besides maximizing its artificial short-run average profit, the firm is actually also maximizing its true long-run average profit. In order for the optimal decision parameters to be the same in the two models, the firm must use the inferred value of b in the PB model. If it makes an error and uses a different value of b, then its long-run average profit will fall short of its potential maximum value.

With the above discussion in mind, we can see that in order for the optimal fill rate F^* in the PB model, which given in column 2 of Table 2-2, to be equal to the optimal fill rate F'^* in the PD model, which is given in column 2 of Table 2-6, b must satisfy

$$b = h\frac{F'^*}{1-F'^*}, \quad \text{in cases 1-3} \tag{2.5}$$

$$b = h\frac{F'^{*2}}{1-F'^{*2}}, \quad \text{in case 4} \tag{2.6}$$

The above expressions give the inferred value of b in the PB model. These expressions imply that if F'^* = 0, then the inferred value of b in the respective cases of the PB model is zero. They also imply that if F'^* = 1, then the inferred value of b in the respective cases of the PB model is infinite. Finally, if F'^* is anywhere between zero and one, then the inferred value of b in the respective case of the PB model is finite. The exact inferred value of b for all the cases of the PB model is shown in the second to last column of Table 2-6. From that column, it can be seen that in cases 1 and 4, the inferred value of b is either zero or infinity, because in these cases F'^* is always equal to zero or one, as was shown in the previous section. In cases 2 and 3, on the other hand, the inferred value of b is either a finite number or infinity, because in these cases F'^* is always either between zero and one, or equal to one, as was also shown in the previous section. Since, in the cases where F'^* is between zero and one, F'^* assumes a continuum of values, depending one the model parameters, as was mentioned in the previous section, the respective inferred value of b also assumes a continuum of values, in these cases.

From column 3 of Table 2-6, it can be seen that in the subcase of case 1 of the PD model, where $p\sqrt{A}B/(1+B) \geq \sqrt{2hk}$, as well as in case 3, the optimal order quantity Q'^* is a function of $D'(F^*)$. From the last column of the same table, it can also be seen that in the respective cases of the PB model, if we use the inferred value of b, shown in the second to last column of Table 2-6, then the resulting optimal order quantity Q^* is given by the same function, but with $D'(F'^*)$ in the place of D. At first, this seems to suggest that in these cases, the inferred value of b, which by definition guarantees that $F^* = F'^*$, does not guarantee that $Q^* = Q'^*$. This further suggests that in these cases, there exist no two models – a PB and a respective PD model – with the same optimal decision parameters. This is true in the short run. However, as was mentioned earlier in this section, if the firm uses the optimal parameters $F^* = F'^*$ and Q^* in the PB model, then as time passes, no matter what the initial value of D is, the average demand rate will drift towards $D'(F^*)$ which is equal to $D'(F'^*)$, assuming that the fill rate F^* is kept constant and equal to F'^*, so that in the long run, its average value will be equal to $D'(F'^*)$. Therefore, in the long run, the optimal order quantity Q^* in the PB model will be equal to the optimal order quantity in the respective PD model.

2.4 Numerical Example

To illustrate the analytical results presented in the previous sections, consider an instance of the classical PB model (case 1) with parameters p = 3, k = 200, h = 1, and D = 100. Consider also the respective classical PD model (case 1) with parameters A = 144 and B = 2, and the same parameters p, k, and h as these in the PB model. From equation (2.2), the maximum and minimum long-run average demand rates in the PD model are $D'(1)$ = 144/1 = 144 and $D'(0)$ = 144/(1 + 2) = 48. This means that the demand rate drops to 1/3 of its maximum value, A, as F' drops from 1 to 0. From column 2 of Table 2-6, the optimal fill rate in the PD model is F'^* = 1, since $p\sqrt{AB}/(B+1) = 3\cdot\sqrt{144}\cdot 2/(1+2) = 24$ is larger than $\sqrt{2hk} = \sqrt{2\cdot 1\cdot 200} = 20$. Moreover, from column 3 of the same table, the optimal order quantity in the PD model is equal to $Q'^* = \sqrt{2\cdot 200\cdot 144/1} = 240$. In order for the optimal fill rate of the PB model, F^*, to be equal to F'^*, then, from the second to last column of Table 2-6, the inferred backorder penalty cost coefficient in the PB model must be equal to $b = \infty$. The last column of Table 2-6 then implies that the optimal order quantity in the classical PB model (case 1) is equal to $Q^* = \sqrt{2\cdot 200\cdot 100/1} = 200$. Clearly, the optimal order quantities in the two models are not the same. However, if the firm uses the optimal decision variables F^* = 1 and $Q^* = \sqrt{2kD/h}$ in the PB model, the average demand rate D will drift from 100 towards $D'(1)$ = 144, so that in the long run, the optimal order quantities in the two models will be equal to each other.

Now, suppose that there is a minimum order quantity $Q_{\min}$ = 1000, instead of the fixed order cost. Then, from our analysis in Sections 2.1 and 2.2, the optimal order quantities in the PB model and the respective PD model with minimum order quantity (case 2) are both equal to $Q_{\min}$ = 1000. From column 2 of Table 2-6, the optimal fill rate in the PD model with minimum order quantity is $F'^* = F_2'$, since pAB = 3·144·2 = 864 is less than $hQ_{\min}$ = 1·1000 = 1000. F_2' is numerically found to be equal to 0.112141. In order for the optimal fill rate of the PB model, F^*, to be equal to F'^*, then, from the second to last column of Table 2-6, the inferred backorder penalty cost coefficient in the

PB model must be equal to b = 0.112141/(1 – 0.112141) = 0.126304. Now, suppose that the minimum order quantity drops to $Q_{\min}$ = 600. Then, from column 2 of Table 2-6, the optimal fill rate in the PD model with minimum order quantity is still $F'^{*} = F_2'$, because pAB = 3·144·2 = 864 is greater than $hQ_{\min}$ = 1·600 = 600, B = 2 is greater than 0.5, and $P'^{*}(F_2')$, which is equal to 154.23, is greater than $pA - hQ_{\min}/2$ = 3·144 - 1·600/2 = 132, where F_2' is numerically found to be equal to 0.219584. If the minimum order quantity drops further to $Q_{\min}$ = 550, however, then from column 2 of Table 2-6, the optimal fill rate in the PD model with minimum order quantity jumps to F'^{*} = 1, because now pAB = 3·144·2 = 864 is still greater than $hQ_{\min}$ = 1·550 = 550, B = 2 is still greater than 0.5, but $P'^{*}(F_2')$, which is equal to 155.613, is less than $pA - hQ_{\min}/2$ = 3·144 – 1·550/2 = 157, where F_2' is numerically found to be equal to 0.252255.

Next, suppose that instead of the minimum order quantity, there is a minimum interorder time $T_{\min}$ = 4. Then, from column 2 of Table 2-6, the optimal fill rate in the PD model with minimum interorder time (case 3) is $F'^{*} = F_3'$, since $2pB/(2 + B)$ = 3·2·2/(2 + 2) = 3 is less than $hT_{\min}$ = 1·4 = 4. From (2.4), F_3' is equal to 0.633975. In order for the optimal fill rate of the PB model, F^{*}, to be equal to F'^{*}, then, from the second to last column of Table 2-6, the inferred backorder penalty cost coefficient in the PB model must be equal to b = 0.633975/(1 – 0.633975) = 1.73205. As was mentioned in Section 2.3, in the short run, the optimal order quantity in the PB model with minimum interorder quantity Q^{*}, which is equal to $DT_{\min}$ = 100·4 = 400, is different from the optimal order quantity in the corresponding PD model, Q'^{*}, which is equal to $D(F'^{*})T_{\min}$ = 144·4/[1 + (1 – 0.633975)·2] = 332.554. However, if the firm uses the optimal decision variables $F^{*} = b/(h + b)$, where b is given by (2.5), and $Q^{*} = DT_{\min}$, in the PB model, the average demand rate D will drift towards $D'(F'^{*})$, and hence the optimal order quantity will drift towards $D(F'^{*})T_{\min}$. Therefore, in the long run, the optimal order quantity in the PB model will be equal to the optimal order quantity in the PD model. Now suppose that the minimum interorder time drops to $T_{\min}$ = 2. Then, from column 2 of Table 2-6, the optimal fill rate in the PD model with minimum interorder time is F'^{*} = 1, because $2pB/(2 + B)$ = 3·2·2/(2 + 2) = 3 is greater than $hT_{\min}$ = 1·2 = 2.

Finally, suppose that instead of the minimum interorder time, there is a minimum starting inventory $I_{\min}$ = 500. Then, from column 2 of Table 2-6, the optimal fill rate in the PD model with minimum starting inventory (case 4) is F'^{*} = 1, since $2pAB/(1 + B)$ = 3·2·144·2/(1 + 2) = 576 is greater than $hI_{\min}$ = 1·500 = 500. In order for the optimal fill rate of the PB model minimum starting inventory, F^*, to be equal to F'^*, then, from (2.6), the inferred backorder penalty cost coefficient in the PB model must be equal to $b = \infty$. Moreover, from Table 2-6, the optimal order quantities in the PB model and the PD with minimum order quantity are both equal to $I_{\min} = 500$.

2.5 Conclusions

The work in this chapter was motivated by our desire to find a credible answer to the question, what could the backorder penalty cost coefficient *b* be? To this end, we proposed to infer the value of *b* for the PB model by connecting *b* to the loss in the long-run average demand rate which is affected by backorders according to Schwartz's PD model. We applied this procedure to the classical PB model and three variations of it in which we replaced the explicit fixed order cost with a constraint on the order quantity, the interorder time, and the starting inventory in each cycle, respectively. We found that for the classical PB model and the variation of the PB model with the minimum starting inventory in each cycle, the optimal fill rate is always either one or zero, which implies that the inferred backorder penalty cost is either infinite or zero, respectively. In the former case, the optimal order quantity is finite, whereas in the latter case it is infinite. For the other two variations, the optimal fill rate is always either one or a finite number between zero and one, which implies that the inferred backorder penalty cost is either infinite or a positive finite number which depends on the model parameters, respectively. In both cases, the optimal order quantity is finite.

Future research could be directed toward repeating this procedure for other PD models, for example models that assume that the long-run average demand rate is either a different function of the long-run average fill rate than the one given by equation (2.2), or a function of some other customer service related performance measure, such as the long-run average backorder waiting time or number of backorders.

Some such functions have been proposed in the literature. For example, Ernst and Cohen (1992) proposed a perturbed demand rate which is a linear function of the fill rate. Using our notation, their function can be written as

$$D'(F') = A\left[1 - B\left(1 - F'\right)\right]$$

where *A* is the maximum potential demand rate corresponding to a fill rate equal to one and *B* is a percentage.

Zipkin (2000) (problem 3.11, p. 69) proposed the perturbed demand rate function

$$D'(W') = \frac{a}{\left[pf(W')\right]^{\beta}}$$

where W' is the average waiting time, α and β are positive constants with $\beta > 1$, and $f(\cdot)$ is an increasing function with $f(0) = 1$. Given that the average waiting time can be expressed as a function of Q' and F' as well as the demand rate itself, namely

$$W' = \frac{Q'(1-F')^2}{2D'(W')}$$

if we substitute W' from the equation above into $D'(W')$, we can see that the perturbed demand rate is a rather complicated function of Q' and F' satisfying

$$D'(Q',F') = \frac{a}{\left[pf(Q'(1-F')^2 / 2D'(Q',F'))\right]^{\beta}}$$

A less complicated alternative would be to replace the average waiting time W' with the average number of backorders, say R', in Zipkin's perturbed demand rate function, i.e., assume that

$$D'(R') = \frac{a}{\left[pf(R')\right]^{\beta}}$$

Given that the average number of backorders R' can be expressed as a function of Q' and F' as follows,

$$R' = \frac{Q'(1-F')^2}{2}$$

then $D'(R')$ can be rewritten as a function of Q' and F' as follows:

$$D'(Q',F') = \frac{a}{\left[pf(Q'(1-F')^2 / 2)\right]^{\beta}}$$

In all the models above, the parameters of the perturbed demand function have to be estimated. As was mentioned in Section 2.2, Schwartz (1966) proposed a procedure for measuring parameters A and B in his model from observed demand data. In general, this is not an easy task; however, it is a better defined task that picking a value for b. Of course, a broader question is, which perturbed demand model is correct? To answer this question, one would have to try different models and use statistical analysis of real demand data to identify the most appropriate model. Finally, two other worthwhile directions for future research following this work would be to include direct backorder costs besides the indirect loss-of-customer-goodwill costs, to examine models with lost sales instead of order backlogging, and to extend this analysis to stochastic inventory models.

Appendix

To solve the four optimization (maximization) problems corresponding to the four cases of the PD model shown in Table 2-4 of Section 2.2, we need to solve the first-order and, if necessary, the second-order optimality conditions, and consider the possibility of obtaining a maximum at the end points of the constraints. To solve the optimality conditions, we use Descartes's rule of signs. This rule, which was first published by Renée Descartes in 1637, states that if the terms of a polynomial $f(x)$ are written in a customary fashion – that is with the terms given in decreasing order of the exponent of x – then the number of positive real roots of the polynomial is either equal to the number of sign changes in the coefficients of successive terms of $f(x)$ or is less than that number by an even number (until 1 or 0 is reached). If any coefficients are zero, they are simply ignored. Similarly, the number of negative real roots of the polynomial is either equal to the number of sign changes in the coefficients of successive terms of $f(-x)$ or is less than that number by an even number (until 1 or 0 is reached) (e.g., see Young and Gregory, 1973).

Solution of the classical PD model (case 1 of Table 2-4)

For the classical PD model, in order to find the optimal order quantity as a function of F', $Q'^*(F')$, we set the first partial derivative of $P'(Q', F')$ with respect to Q' equal to zero and solve the resulting equation. This equation is quadratic in Q' and has two solutions, one of which is negative. The only positive and therefore acceptable solution is

$$Q'^*(F') = \sqrt{\frac{2kD'(F')}{hF'^2}} = \sqrt{\frac{2kA}{hF'^2\left[1+(1-F')B\right]}} \tag{2.7}$$

Let $P'^*(F')$ be the average profit as a function of F' when the optimal order quantity is used, i.e.,

$$P'^*(F') = P'(Q'^*(F'), F') = pD'(F') - F\sqrt{2khD'(F')} = \frac{pA}{1+(1-F')B} - F\sqrt{\frac{2khA}{1+(1-F')B}} \tag{2.8}$$

To find the optimal fill rate, F'^*, we set the first derivative of $P'^*(F')$ equal to zero, solve the resulting equation, and examine the values of the average profit and its derivative at the end points of the interval [0, 1].

The first derivative of the average profit $P'^*(F')$ is

$$\frac{dP'^*(F')}{dF'} = \frac{pAB}{\left[1+(1-F')B\right]^2} - \frac{\sqrt{2khA}\left[2+(2-F')B\right]}{2\sqrt{1+(1-F')B}} \tag{2.9}$$

Setting the first derivative of $P'^*(F')$ equal to zero, performing a change of variables from F' to Y, where $Y = 1 + (1 - F')B$, and rearranging terms, yields the following 5th degree polynomial equation in Y,

$$Y^5 + 2(1+B)Y^4 + (1+B)^2 Y^3 - \frac{2Ap^2B^2}{hk} = 0 \tag{2.10}$$

According to Descartes's rule of signs, the polynomial on the lhs of equation (2.10) has exactly one positive real root and exactly two or zero negative real roots. For each real root, Y_n, there corresponds a real root, F'_n, of the rhs of expression (2.9), which is given by $F'_n = 1 - (Y_n - 1)/B$. Since F' represents the long-run, average fill rate, it must take values in the interval [0, 1]. Note that if $Y_n < 1$, then $F'_n > 1$, whereas if $Y_n > 1 + B$, then $F'_n < 0$. This implies that for each negative real root, Y_n, if there are any, the corresponding root F'_n is greater than one. It also implies that the root F'_n corresponding to the only positive real root, Y_n, lies in the interval [0, 1] if and only if $Y_n \in [1, 1+B]$. This means that there is at most one real root of the rhs of equation (2.9) that may lie in the interval [0, 1].

With the above result in mind, to find the optimal fill rate, F'^*, we proceed by examining the average profit and its derivative at the end points, 0 and 1. From (2.9), it is easy to see that the first derivative of the average profit at the two end points, 0 and 1, is given respectively by

$$\left.\frac{dP'^*(F')}{dF'}\right|_{F'=0} = \frac{pAB - (1+B)\sqrt{1+B}\sqrt{2khA}}{(1+B)^2} \tag{2.11}$$

$$\left.\frac{dP'^*(F')}{dF'}\right|_{F'=1} = pAB - \sqrt{2khA}\left(1 + \frac{B}{2}\right) \tag{2.12}$$

From (2.8) it is also easy to see that the average profit at the end points, 0 and 1 is given respectively by

$$P'^*(0) = \frac{pA}{1+B} \tag{2.13}$$

$$P'^*(1) = pA - \sqrt{2khA} \tag{2.14}$$

Now, suppose that $P'^*(0) > P'^*(1)$, which, from (2.13) and (2.14), is true if and only if $pAB < (B+1)\sqrt{2hkA}$. The latter condition, which can be rewritten as $p\sqrt{A}B/(1+B) < \sqrt{2hk}$, implies that the first derivative of the average profit at $F' = 0$, which is given by (2.11), is always negative. This means that as F' increases starting from zero, the average profit, which starts at $P'^*(0)$, either continuously decreases in the interval [0, 1], or continuously decreases until it reaches a minimum at the only real root of the rhs of expression (2.9) which may possibly lie in the interval [0, 1], and then continuously increases – since there is at most one real root in the interval [0, 1] – until it reaches $P'^*(1)$ at $F' = 1$. Given our initial assumption that $P'^*(0) > P'^*(1)$, this further implies that the maximum average profit in the interval [0, 1] is attained at $F' = 0$.

Now, suppose that $P'^*(0) < P'^*(1)$, which, from (2.13) and (2.14), is true if and only if $p\sqrt{A}B/(1+B) > \sqrt{2hk}$. Then, the first derivative of the average profit at $F' = 1$, which is given by (2.12), is always positive. This means that as F' decreases starting from one, the average profit, which starts at $P'^*(1)$, either continuously decreases in the interval [0, 1], or continuously decreases until it reaches a minimum at the only real root of expression (2.9) which may possibly lie in the interval [0, 1], and then continuously increases – since there is at most one real root in the interval [0, 1] – until it reaches $P'^*(0)$ at $F' = 0$. Given our initial assumption that $P'^*(0) < P'^*(1)$, this further implies that the maximum average profit in the interval [0, 1] is attained at $F' = 1$.

Following the same argument, it can also be shown that if we assume that $P'^*(0) = P'^*(1)$, which, from (2.13) and (2.14), is true if and only if $p\sqrt{A}B/(1+B) = \sqrt{2hk}$, then the maximum average profit in the interval [0, 1] is attained at both $F' = 0$ and $F' = 1$.

Solution of the PD model with a minimum order quantity (case 2 of Table 2-4)

For the PD model with a minimum order quantity $Q_{\min}$, in order to find the optimal order quantity Q'^* note that the average profit function $P'(Q', F')$ is decreasing in Q'; therefore, the optimal order quantity, Q'^*, should be as small as possible as long as the minimum order quantity constraint is not violated. This means that $Q'^* = Q_{\min}$.

Let $P'^*(F')$ be the average profit as a function of F' when the optimal order quantity is used, i.e.,

$$P'^*(F') = P'(Q'^*, F') = P'(Q_{\min}, F') = pD'(F') - h\frac{Q_{\min}F'^2}{2} = \frac{pA}{1+\left(1-F'\right)B} - h\frac{Q_{\min}F'^2}{2} \quad (2.15)$$

To find the optimal fill rate, F'^*, we set the first derivative of $P'^*(F')$ equal to zero, solve the resulting equation, and examine the values of the average profit and its derivative at the end points of the interval [0, 1].

The first derivative of the average profit $P'^*(F')$, given by (2.15), is given by (2.3), i.e.,

$$\frac{dP'^*(F')}{dF'} = \frac{pAB}{\left[1+(1-F')B\right]^2} - hQ_{\min}F' \quad (2.16)$$

The above expression implies that the first derivative of the average profit at $F' = 0$ is always positive. Setting the above expression equal to zero and rearranging terms yields the following cubic equation in F':

$$hB^2Q_{\min}F'^3 - (2hBQ_{\min})(1+B)F'^2 + hQ_{\min}(1+B)^2 F' - pBA = 0 \tag{2.17}$$

According to Descartes's rule of signs, the lhs of the above equation has exactly three or one real positive roots, and no negative real roots. To further investigate how many of the positive real roots lie in the interval [0, 1], we perform a change of variables from F' to Y, where $Y = 1 - F'$, and reset expression (2.16) equal to zero. After rearranging terms we obtain the following cubic equation in Y:

$$hB^2Q_{\min}Y^3 + hBQ_{\min}(2-B)Y^2 + 2hQ_{\min}(0.5-B)Y + pBA - hQ = 0 \tag{2.18}$$

For each real root, Y_n, of the cubic polynomial on the lhs of the above equation, there corresponds a real root of the lhs of equation (2.17), F'_n, which is given by $F'_n = 1 - Y_n$. To determine how many of the roots F'_n lie in [0, 1], we proceed as follows.

First, suppose that $pBA < hQ_{\min}$. Then, according to Descartes's rule of signs, it can be easily shown that the lhs of the equation (2.18) has exactly one positive real root and two or zero negative real roots, regardless of the value of B. This is done by examining the cases where B is less than 0.5, B is between 0.5 and 2, and B is greater than 2. This implies that for each negative real root Y_n, if any, the corresponding root F'_n is greater than one. It also implies that the root F'_n corresponding to the only positive real root Y_n is less than one. Given than the lhs of equation (2.17) has no negative real roots, as was mentioned above, this further implies that the root F'_n corresponding to the only positive real root Y_n is also greater than zero. To summarize, if $pBA < hQ_{\min}$, the cubic polynomial on the lhs of equation (2.17) always has exactly on real root, say F'_R, in the interval [0, 1] and two or zero real roots which are greater than one. Moreover, given that the first derivative of the average profit at $F' = 0$ is always positive, as was mentioned above, then as F' increases starting from zero, the average profit continuously increases in the interval $[0, F'_R)$, reaches a maximum at F'_R, and decreases in the interval $(F'_R, 1]$, since there are no other real roots in the interval [0, 1]. This also means that the first derivative of the average profit at $F' = 1$ is negative, which from (2.16) is true if and only if $pAB < hQ_{\min}$. The latter condition coincides with our original assumption.

Next, suppose that $pBA > hQ_{\min}$ and $B < 0.5$. Then, according to Descartes's rule of signs, the lhs of the equation (2.18) has no positive real roots and exactly three or one negative real roots. This implies that for each negative real root Y_n the corresponding real root F'_n is greater than one. In other words, there are no real roots F'_n that lie in the interval [0, 1]. Given that the first derivative of the average profit at $F' = 0$ is always positive, as was mentioned above, this further implies that as

F' increases starting from zero, the average profit continuously increases in the interval [0, 1] – since there are no real roots in the interval [0, 1] – reaching a maximum at $F' = 1$. This also means that the first derivative of the average profit at $F' = 1$ is positive, which from (2.16) is true if and only if $pAB > hQ_{\min}$. The latter condition coincides with our original assumption.

Finally, suppose that $pBA > hQ_{\min}$ and $B > 0.5$. Then, according to Descartes's rule of signs, the lhs of the equation (2.18) has exactly two or zero positive real roots and one negative real root. This implies that the real root F'_n corresponding to the only negative real root Y_n is greater than one. It also implies that for each positive real root Y_n, if any, the corresponding real root F'_n is less than one. Given than the lhs of equation (2.17) has no negative real roots, as was mentioned above, this further implies that the roots F'_n corresponding to the positive real roots Y_n, if any, are also greater than zero. To summarize, if $pBA > hQ_{\min}$ and $B > 0.5$, the cubic polynomial on the lhs of equation (2.17) always has two or zero real roots in the interval [0, 1] and one real root which is greater than one. If there are two real roots in the interval [0, 1], then, given that the first derivative of the average profit at $F' = 0$ is always positive, it is straightforward to see that the smallest root, say F'_2, yields a local maximum of the average profit and the second root yields a local minimum. In this case, if $P'^*(F'_2) > P'^*(1)$, which from (2.15), can be rewritten as $P'^*(F'_2) > pA - hQ_{\min}/2$, then the average profit is maximized at $F' = F'_2$ in the interval [0, 1]; otherwise, it is maximized at $F' = 1$. On the other hand, if there are no real roots in the interval [0, 1], then, given that the first derivative of the average profit at $F' = 0$ is always positive, it is straightforward to see that the average profit is maximized at $F' = 1$ in the interval [0, 1].

The above analysis was extended to include the cases where $pBA = hQ_{\min}$ and/or $B = 0.5$. We omit the details here for space considerations.

Solution of the PD model with a minimum interorder time (case 3 of Table 2-4)

For the PD model with a minimum interorder time $T_{\min}$, in order to find the optimal order quantity Q'^* note that the average profit function $P'(Q', F')$ is decreasing in Q', so the optimal order quantity, Q'^*, should be as small as possible as long as the minimum order quantity constraint is not violated. This means that $Q'^* = D(F')T_{\min}$.

Let $P'^*(F')$ be the average profit as a function of F' when the optimal order quantity is used, i.e.,

$$P'^{*}(F') = P'(Q'^{*}, F') = P'(D'(F')T_{\min}, F') = pD'(F') - h\frac{D'(F')T_{\min}F'^{2}}{2}$$
$$= \frac{pA}{1+\left(1-F'\right)B} - \frac{hAT_{\min}F'^{2}}{2\left[1+\left(1-F'\right)B\right]} \tag{2.19}$$

To find the optimal fill rate, F'^{*}, we set the first derivative of $P'^{*}(F')$ equal to zero, solve the resulting equation, and examine the values of the average profit and its derivative at the end points of the interval [0, 1].

The first derivative of the average profit $P'^{*}(F')$, given by (2.19), is

$$\frac{dP'^{*}(F')}{dF'} = \frac{pAB}{\left[1+(1-F')B\right]^{2}} - \frac{hAT_{\min}F'}{1+(1-F')B} - \frac{hABT_{\min}F'^{2}}{2\left[1+(1-F')B\right]^{2}} \tag{2.20}$$

The above expression implies that the first derivative of the average profit at $F' = 0$ is always positive. Setting the above expression equal to zero and rearranging terms yields the following quadratic equation in F':

$$hABT_{\min}F'^{2} - \left(2hAT_{\min}\right)\left(1+B\right)F' + 2pAB = 0 \tag{2.21}$$

The above equation has the following two solutions:

$$1 + \frac{1}{B} \pm \sqrt{\frac{\left(1+B\right)^{2}}{B^{2}} - \frac{2p}{h^{2}T_{\min}^{2}}} \tag{2.22}$$

If the term under the square root in the above expression is negative, then both solutions are complex numbers. In order for the term under the square root to be negative, the following condition must hold:

$$2pB^{2} > h\left(1+B\right)^{2}T_{\min}$$

Suppose that the above condition does hold. Then equation (2.21) has no real solutions. Given that the first derivative of the average profit at $F' = 0$ is always positive, as was mentioned above, then as F' increases starting from zero, the average profit continuously increases since equation (2.21) has no real solutions; therefore, the maximum average profit for values of F' in the interval [0, 1] is attained at $F' = 1$.

If the term under the square root in expression (2.22) is positive, then both solutions of the quadratic equation (2.21) are real numbers; however, one of them is always greater than one. The only real solution that may lie in the interval [0, 1] is the solution F'_{3} given by (2.4). Given that the first derivative of the average profit at $F' = 0$ is always positive, then in order for the above solution to lie in the interval [0, 1], the first derivative of the average profit at $F' = 1$ must be negative. From

(2.20), the latter is true if and only if $2pB < h(2 + B)T_{\min}$, which can be rewritten as $2pB/(2 + B) < hT_{\min}$. If this condition holds, then the maximum average profit for values of F' in the interval [0, 1] is attained at $F' = F'_3$; otherwise it is attained at $F' = 1$.

Solution of the PD model with a minimum starting inventory (case 4 of Table 2-4)

For the PD model with a minimum starting inventory $I_{\min}$, in order to find the optimal order quantity Q'^* note that the average profit function $P'(Q', F')$ is decreasing in Q', so the optimal order quantity, Q'^*, should be as small as possible as long as the minimum order quantity constraint is not violated. This means that $Q'^*(F') = I_{\min}/F'$.

Let $P'^*(F')$ be the average profit as a function of F' when the optimal order quantity is used, i.e.,

$$P'^*(F') = P'(Q'^*, F') = P'(I_{\min}/F', F') = pD'(F') - h\frac{I_{\min}F'}{2} = \frac{pA}{1+(1-F')B} - h\frac{I_{\min}F'}{2} \quad (2.23)$$

To find the optimal fill rate, F'^*, we examine the first and second derivative of $P'^*(F')$, as well as the values of $P'^*(F')$ at the end points of the interval [0, 1].

The first and second derivative of the average profit $P'^*(F')$, given by (2.23), are

$$\frac{dP'^*(F')}{dF'} = \frac{pAB}{\left[1+(1-F')B\right]^2} - \frac{hI_{\min}}{2} \quad (2.24)$$

$$\frac{dP'^{*2}(F')}{dF'^2} = \frac{2pAB^2}{\left[1+(1-F')B\right]^3} \quad (2.25)$$

From equation (2.25), it is obvious that for every $F' \in [0, 1]$, the second derivative of $P'^*(F')$ is always positive. This means that the average profit is convex in F' in the interval [0, 1]; therefore, the optimal fill rate, F'^*, coincides with one of the two end points, 0 or 1. More specifically, if $P'^*(0) > P'^*(1)$, which from (2.23) is true if and only if $2pAB/(1 + B) < hI_{\min}$, then $F'^* = 0$. Conversely, if $P'^*(0) < P'^*(1)$, which from (2.23) is true if and only if $2pAB/(1 + B) > hI_{\min}$, then $F'^* = 1$. Finally, if $P'^*(0) = P'^*(1)$, which from (2.23) is true if and only if $2pAB/(1 + B) = hI_{\min}$, both 0 and 1 are optimal.

Chapter 3 Competing for Customer Goodwill on Product Availability

In this chapter, we develop a newsvendor-type model of two suppliers that compete to sell the same type of items to a customer, repetitively, in discrete periods, for an infinite time horizon. At the beginning of each period, each supplier orders a number of items that are delivered to him immediately. In each period, the customer randomly chooses one of the two suppliers and demands from him a random number of items. The probability of choosing a supplier depends on the so-called “credibility level” of this supplier, which reflects the customer's estimate of the supplier's relative credibility based on the history of service – measured in terms of product availability – that both suppliers have provided to the customer in the past. The credibility levels of the suppliers change dynamically based on the quality of service – good or poor – that the customer receives in each period.

In Section 3.1, we formulate the problem of finding optimal stationary ordering policies for both suppliers at equilibrium as a stochastic dynamic game. In Section 3.2, we propose a numerical solution technique for solving the resulting optimality conditions, and in Section 3.3, we implement this technique for several instances of the problem. In all instances, the optimal ordering policy for each supplier turns out to be an order-up-to policy. In Section 3.4, we restrict our attention to the case where each supplier has only two credibility levels, a low and a high, such that, when in the low level, he is never chosen by the customer, and when in the high level, he is always chosen by the customer. For this case, we assume that each supplier uses a credibility level-dependent order-up-to policy. This leads to a Markov Decision Process with two decision makers. We numerically solve the resulting optimality conditions at equilibrium to find the optimal order-up-to levels of both suppliers. Finally, we conclude in Section 3.5.

3.1 Model Description and Mathematical Formulation

We consider a model of two suppliers that compete to sell the same type of items to a single customer. Their competition takes places repetitively, in discrete time periods, for an infinite time horizon. We make the following assumptions.

(A1). *In each period, the customer randomly chooses one of the two suppliers and demands from him a random number of items.*

Let w^t be the customer's demand in period *t*.

(A2). *The demands* $w^0, w^1, \ldots,$ *are i.i.d. discrete random variables with probability mass function* $p(\cdot)$*, cumulative distribution function* $T(\cdot)$*, and mean* θ*.*

(A3). *The probability with which the customer chooses a supplier depends only on the so-called "credibility level" of the supplier, which reflects the customer's estimate of the supplier's relative credibility based on the history of service (measured in terms of product availability) that both suppliers have provided to the customer.*

Assumption (A3) states that the suppliers compete for the customer's goodwill based only on the history of product availability. This implies that all other competition drivers, such as price, after-sales service, etc, are more or less the same for both suppliers.

Let a^t be the credibility level of supplier 1 at the beginning of period *t*.

(A4). a^t *may belong to a number of discrete states,* $0,1,\ldots,M$.

(A5). *The sum of the credibility levels of both suppliers is constant and equal to M at all periods; hence, the credibility level of supplier 2 at the beginning of period t is* $M - a^t$.

Assumption (A5) implies that the credibility level of one supplier is complementary and therefore relative to that of the other supplier. This is a reasonable assumption if the customer (a) has no other option but to buy the items he demands from one of the two suppliers, i.e., the two suppliers form a pure duopoly, and (b) will not change the distribution (e.g., the mean) of his demand even if he repeatedly receives poor service from both suppliers; i.e., the customer absolutely needs the items he demands.

(A6). *If, in period t, supplier 1 is chosen by the customer and is able to meet all the demand (good service), or if supplier 2 is chosen by the customer and is unable to meet all the demand* *(poor service), then, at the beginning of period $t+1$, the credibility level of supplier 1 increases by one, i.e. $a^{t+1} = a^t + 1$ (therefore, the credibility level of supplier 2 decreases by one), unless $a^t = M$, in which case $a^{t+1} = M$. The opposite is true if supplier 2 is chosen by the customer.*

Assumption (A6) describes the dynamic evolution of a^t. It implies that the customer's response to good service from one supplier is exactly the same as his response to poor service from the other supplier. This assumption simplifies the analysis, because it renders $\{a^t, t = 0,1,\ldots\}$ a birth-death process. Although it appears to be restrictive, it is not crucial in the sense that it does not result in any loss of generality. Had we assumed that the customer's response to good and to poor service is asymmetric, the evolution of a^t would have been more complicated, but the structure, and hence the analysis, of the model would remain essentially unchanged. A simple example of asymmetric behavior would be the case where, if a supplier meets the demand, his credibility level increases by one, whereas if he does not meet all the demand, his credibility level decreases by two.

Let $q_i(a)$ be the probability that the customer chooses supplier *i* in a period, given that supplier 1's credibility level is a at the beginning of that period. Assumption (A5) implies that $q_1(a) + q_2(a) = 1,\ a = 0,1,\ldots,M$; therefore $\overline{q}_1(a) \equiv 1 - q_1(a) = q_2(a)$ and $\overline{q}_2(a) \equiv 1 - q_2(a) = q_1(a)$.

(A7). $q_1(a) \geq q_1(a')$ and $q_2(a) \leq q_2(a'),\ a > a'$.

Assumption (A7) implies that the probability with which the customer chooses a supplier is non-decreasing in the supplier's credibility level. In general, we would expect $q_1(a)$ and $q_2(M-a)$ to be similar in shape if the customer's behavior towards both suppliers is symmetric, at least qualitatively, if not quantitatively. A simple assumption would be that $q_1(a)$ is linear in *a*, e.g. $q_1(a) = q_1^{\min} + a\left(1 - q_1^{\min} - q_2^{\min}\right)\big/M$, where $q_i^{\min}$ is the smallest probability of choosing supplier *i* when his credibility level is at its lowest possible value. In general, the shape of $q_1(a)$ should depend on the customer's response to good and to poor service from the same supplier. For example, if the customer is more reluctant to significantly change the probability of choosing supplier 1, if supplier 1's credibility level is already too low or too high, than if it is medium, then it might be reasonable to assume that $q_1(a)$ is "*S*"-shaped, being flatter towards the ends than around the middle. If the customer is willing to "forgive" but not forget one or more or more poor services in a row by supplier, then it might be reasonable to assume that $q_1(a)$ is piece-wise constant in *a*. The model is flexible and can accommodate different customer behaviors.

(A8). *At the very beginning of each period, each supplier orders a number of items that are delivered to him immediately. When ordering, he has complete information about his as well as his competitor's current inventory surplus/backlog level and credibility level, but he has no knowledge of his competitors ordering decision.*

Assumption (A8) is perhaps the most limiting assumption, because in practice it is unlikely that one supplier has complete information of the other supplier's state. In fact, if a supplier has no information about the other supplier's inventory surplus/backlog level, he can not know with certainty his own credibility state. It is natural, however, to look into the complete-information case first, before tackling the more complicated incomplete-information case.

Let x_i^t be the inventory surplus/backlog of supplier *i* at the very beginning of period *t*; x_i^t may take positive or negative values. Let u_i^t be the replenishment quantity ordered (and immediately delivered) by supplier *i* at the beginning of period *t*.

(A9). $u_i^t \ge \left(-x_i^t\right)^+$, *where* $\left(x\right)^+ \equiv \max\left(0, x\right)$.

Assumption (A9) implies that if the supplier chosen by the customer in a period is unable to meet all the demand, then the unmet demand is backordered with this supplier, who must satisfy it at the beginning of the next period. This means that the customer does not switch suppliers within each period. This assumption is reasonable if the customer routinely demands items (e.g. consumables) in each period, without first checking about their availability, and is willing to tolerate – albeit, with some dissatisfaction reflected in suppliers' credibility levels – the wait for one period. Assumption (A9) appears somewhat restrictive in that it implies that the supplier's order "must" be big enough to satisfy any backorders from the previous period. If the customer is willing to tolerate the wait for one period at no direct cost to the supplier, however, it is easy to see that it would be anyway optimal for the supplier to cover the previous period's backorders, if any, because this would maximize his profit. In this case, the constraint of Assumption (A9) would be redundant.

Based, on the assumptions above, the inventory surplus/backlog of supplier *i* evolves according to the following stochastic dynamic equation:

$$x_i^{t+1} = f_i(x_i^t, a^t, u_i^t, w^t)$$

where

$$f_i(x_i, a, u_i, w) = \begin{cases} x_i + u_i - w, & \text{with probability } q_i(a) \\ x_i + u_i, & \text{with probability } \overline{q}_i(a) \end{cases}$$

(A10). *In each period, supplier i receives a reward (selling price)* r_i *per unit for the items he sells and pays a procurement cost* c_i *per unit for the items he orders. He also incurs an inventory holding cost* h_i *per unit for the items he stocks.*

Since, by Assumptions (A1)-(A3), the suppliers do not compete on price, it is reasonable to expect that $r_1 = r_2 \equiv r$. We should also expect that $r > c_i,\ i = 1,2$; otherwise, it makes no sense for the suppliers to sell the items. Finally, it is reasonable to assume that $h_i \ll c_i$, e.g. $h_i = \beta c_i,\ i = 1,2$, where β is the interest rate per period.

Let $g_i(x_i, a, u_i, w)$ be the profit of supplier *i* in a period, as a function of his inventory surplus/backlog at the very beginning of the period, x_i, his order quantity at the beginning of the period, u_i, the credibility level of supplier 1 at the beginning of the period, *a*, and the customer demand in the period, *w*; $g_i(x_i, a, u_i, w)$ is given by the following expression:

$$g_i(x_i,a,u_i,w) = \begin{cases} r_i w - c_i u_i - h_i\left(x_i + u_i - w\right)^+, & \text{with probability } q_i(a) \\ -c_i u_i - h_i\left(x_i + u_i\right) & \text{with probability } \overline{q}_i(a) \end{cases} \quad u_i \geq \left(-x_i\right)^+$$

Let $G_i(x_i, a, u_i)$ be the expected value of $g_i(x_i, a, u_i, w)$ over all possible values of *w*. It is easy to see that

$$\begin{aligned} G_i(x_i,a,u_i) &\equiv \underset{w|x_i,a,u_i}{E}\left[g_i(x_i,a,u_i,w)\right] \\ &= r_i\theta q_i(a) - c_i u_i - h_i\left(\left(x_i+u_i\right)\overline{q}_i(a) + E\left[\left(x_i+u_i-w\right)^+\right]q_i(a)\right) \\ &= r_i\theta q_i(a) - c_i u_i - h_i\left(\left(x_i+u_i\right)\overline{q}_i(a) + \left[\sum_{w=0}^{x_i+u_i}\left(x_i+u_i-w\right)p(w)\right]q_i(a)\right),\ u_i \geq \left(-x_i\right)^+ \end{aligned}$$

Suppose that supplier *i* uses a stationary ordering policy μ_i which maps the system state (x_1, x_2, a) into the control $u_i = \mu_i(x_1, x_2 a)$, and μ_i is such that $\mu_i(x_1, x_2 a) \geq \left(-x_i\right)^+$ for all inventory states x_i, so that μ_i is admissible according to Assumption (A9). Then, under the stationary ordering policies μ_1 and μ_2 of the two suppliers, the state of the system, (x_1^t, x_2^t, a^t), is a discrete-time Markov chain with one-step transition probabilities,

$$\begin{aligned} P_{(x_1,x_2,a)\left(x_1+\mu_1(x_1,x_2,a)-w,\,x_2+\mu_2(x_1,x_2,a),\,\overline{a+1}\right)} &= q_1(a)\,p(w),\ w \leq x_1 + \mu_1(x_1,x_2,a) \\ P_{(x_1,x_2,a)\left(x_1+\mu_1(x_1,x_2,a)-w,\,x_2+\mu_2(x_1,x_2,a),\,\underline{a-1}\right)} &= q_1(a)\,p(w),\ w > x_1 + \mu_1(x_1,x_2,a) \end{aligned}$$

$$P_{(x_1,x_2,a)\left(x_1+\mu_1(x_1,x_2,a),x_2+\mu_2(x_1,x_2,a)-w,\underline{a-1}\right)} = q_2(a)\,p(w),\ w \le x_2+\mu_2(x_1,x_2,a)$$
$$P_{(x_1,x_2,a)\left(x_1+\mu_1(x_1,x_2,a),x_2+\mu_2(x_1,x_2,a)-w,\overline{a+1}\right)} = q_2(a)\,p(w),\ w > x_2+\mu_2(x_1,x_2,a)$$

where $\underline{a} \equiv \max\left(0,a\right) \equiv \left(a\right)^+$ and $\overline{a} \equiv \min(a,M)$.

Suppose that supplier $\overline{i}$ (supplier i's competitor) uses an admissible stationary ordering policy $\mu_{\overline{i}}$. Then the problem of supplier i is to find a stationary ordering policy μ_i that maximizes his long-run average profit,

$$J_i^{\mu_{\overline{i}}} \equiv \max_{\mu_i(x_i^t,x_2^t,a^t)} \lim_{T\to\infty} \frac{1}{T} E\left[\sum_{t=0}^{T-1} g_i(x_i^t,a^t,\mu_i(x_1^t,x_2^t,a^t),w^t \mid \mu_{\overline{i}}(x_1^t,x_2^t,a^t))\right]$$

Using standard dynamic programming arguments, $J_i^{\mu_{\overline{i}}}$ must satisfy the following Bellman's (optimality) equation

$$\begin{aligned} J_i^{\mu_{\overline{i}}} + V_i^{\mu_{\overline{i}}}(x_1,x_2,a) &= (TV_i^{\mu_{\overline{i}}})(x_1,x_2,a) \\ &\equiv \max_{u_i \ge (-x_i)^+} \left\{(TV_{i,u_i}^{\mu_{\overline{i}}})(x_1,x_2,a)\right\},\ \text{for all } (x_1,x_2,a) \end{aligned} \tag{3.1}$$

where $V_i^{\mu_{\overline{i}}}(x_1,x_2,a)$ is the differential profit of supplier i in state (x_1,x_2,a) when his competitor, supplier $\overline{i}$, uses stationary ordering policy $\mu_{\overline{i}}$, and $(TV_{i,u_i}^{\mu_{\overline{i}}})(x_1,x_2,a)$ is a mapping given by the following two expressions for i = 1, 2:

$$\begin{aligned} (TV_{1,u_1}^{\mu_2})(x_1,x_2,a) \equiv G_1(x_1,a,u_1) + \\ q_1(a)\Bigg[\sum_{w=0}^{x_1+u_1} p(w)V_1^{\mu_2}(x_1+u_1-w,x_2+\mu_2(x_1,x_2,a),\overline{a+1}) + \\ \sum_{w=x_1+u_1+1}^{\infty} p(w)V_1^{\mu_2}(x_1+u_1-w,x_2+\mu_2(x_1,x_2,a),\underline{a-1})\Bigg] + \\ q_2(a)\Bigg[\sum_{w=0}^{x_2+\mu_2(x_1,x_2,a)} p(w)V_1^{\mu_2}(x_1+u_1,x_2+\mu_2(x_1,x_2,a)-w,\underline{a-1}) + \\ \sum_{w=x_2+\mu_2(x_1,x_2,a)+1}^{\infty} p(w)V_1^{\mu_2}(x_1+u_1,x_2+\mu_2(x_1,x_2,a)-w,\overline{a+1})\Bigg] \end{aligned}$$

$$
\begin{aligned}
(TV_{2,u_2}^{\mu_1})(x_1, x_2, a) \equiv G_2(x_2, a, u_2) + \\
& q_1(a)\left[\sum_{w=0}^{x_1+\mu_1(x_1,x_2,a)} p(w)V_2^{\mu_1}(x_1+\mu_1(x_1,x_2,a)-w, x_2+u_2, \overline{a+1}) + \right. \\
& \left. \sum_{w=x_1+\mu_1(x_1,x_2,a)+1}^{\infty} p(w)V_2^{\mu_1}(x_1+\mu_1(x_1,x_2,a)-w, x_2+u_2, \underline{a-1})\right] + \\
& q_2(a)\left[\sum_{w=0}^{x_2+u_2} p(w)V_2^{\mu_1}(x_1+\mu_1(x_1,x_2,a), x_2+u_2-w, \underline{a-1}) + \right. \\
& \left. \sum_{w=x_2+u_2+1}^{\infty} p(w)V_2^{\mu_1}(x_1+\mu_1(x_1,x_2,a), x_2+u_2-w, \overline{a+1})\right]
\end{aligned}
$$

Let μ_i^* be the optimal stationary policy of supplier i at equilibrium, i.e. when his competitor also uses his optimal stationary policy, $\mu_{\bar{i}}^*$. Then, the μ_i^*, $i = 1, 2$, must jointly satisfy the optimality conditions

$$
\mu_i^*(x_1, x_2, a) = \arg \max_{u_i \geq (-x_i)^+} (TV_{i,u_i}^{\mu_{\bar{i}}^*})(x_1, x_2, a), \text{ for all } (x_1, x_2, a), \; i = 1, 2 \tag{3.2}
$$

Let $J_i^* \equiv J_i^{\mu_{\bar{i}}^*}$ and $V_i^*(x_1, x_2, a) \equiv V_i^{\mu_{\bar{i}}^*}(x_1, x_2, a)$ be the optimal long-run average profit and differential profit of supplier i in state (x_1, x_2, a) at equilibrium, assuming an equilibrium exists. Then, by (3.1), J_i^* and $V_i^*(x_1, x_2, a)$ must satisfy

$$
J_i^* + V_i^*(x_1, x_2, a) = (TV_i^{\mu_{\bar{i}}^*})(x_1, x_2, a), \text{ for all } (x_1, x_2, a), \; i = 1, 2 \tag{3.3}
$$

3.2 Numerical Solution Technique

To solve equation (3.2), we will use value iteration. Since this method requires that the state space be finite, we will truncate the infinite state-space of (x_1, x_2) to a finite state-space, by imposing the constraints,

$$
x_i^{\min} \leq x_i \leq x_i^{\max}, \; i = 1, 2
$$

for some lower and upper bounds, $x_i^{\min}$ and $x_i^{\max}$. If these bounds are large enough, we expect their influence to be negligible for states (x_1, x_2) away from the bounds.

At each step of the iteration, we will update the values of the optimal control and differential profit functions, $\mu_i^*(x_1, x_2, a)$ and $V_i^*(x_1, x_2, a)$, for all states (x_1, x_2, a), based on the values from the previous step, until they converge. The average profit J_i^* can then be obtained from equation (3.3).

Actually, equation (3.3) has a degree of freedom, because $V_i^*(x_1, x_2, a)$ is defined as a *relative* profit vector. In other words, if we add the same constant to $V_i^*(x_1, x_2, a)$, for all (x_1, x_2, a), equation (3.3) will still hold. To eliminate this degree of freedom, we require that $V_i^*(x_1', x_2', a') = 0$ for some fixed state (x_1', x_2', a'), say $(x_1', x_2', a') = (0,0,0)$. Then from (3.3), $J_i^* = (TV_i^{\mu_{\bar{i}}^*})(x_1', x_2', a'),\ i = 1,2.$

Let $\mu_i^{(n)}(x_1, x_2, a)$ and $V_i^{(n)}(x_1, x_2, a)$ be the approximations of $\mu_i^*(x_1, x_2, a)$ and $V_i^*(x_1, x_2, a)$ at the n^{th} iteration, respectively. Then the approximation of $\mu_i^*(x_1, x_2, a)$ at the $(n+1)^{\text{th}}$ iteration, $\mu_i^{(n+1)}(x_1, x_2, a)$, will be given by another nested value iteration that solves equilibrium condition (3.2). Let $\mu_i^{(n+1)(k)}(x_1, x_2, a)$ be the approximation of $\mu_i^{(n+1)}(x_1, x_2, a)$ at the k^{th} nested iteration, given the $V_i^{(n)}(x_1, x_2, a)$. Then the successive approximation of $\mu_i^{(n+1)}(x_1, x_2, a)$ at the $(k+1)^{\text{th}}$ iteration is given by

$$\begin{aligned} \mu_1^{(n+1)(k+1)}(x_1, x_2, a) &= \arg\max_{u_1 \geq (-x_1)^+} (TV_{1,u_1}^{(n),\mu_2^{(n+1)(k)}})(x_1, x_2, a) \\ \mu_2^{(n+1)(k+1)}(x_1, x_2, a) &= \arg\max_{u_2 \geq (-x_2)^+} (TV_{2,u_2}^{(n),\mu_1^{(n+1)(k+1)}})(x_1, x_2, a) \end{aligned} \tag{3.4}$$

If the above iteration converges to some values $\mu_i^{(n+1)}(x_1, x_2, a),\ i = 1,2,$ then these values must satisfy

$$\mu_i^{(n+1)}(x_1, x_2, a) = \arg\max_{u_i \geq (-x_i)^+} (TV_{i,u_i}^{(n),\mu_{\bar{i}}^{(n+1)}})(x_1, x_2, a),\ i = 1,2 \tag{3.5}$$

Once convergence for the $\mu_i^{(n+1)}(x_1, x_2, a),\ i = 1,2,$ for all states (x_1, x_2, a) is attained, the approximation of $V_i^*(x_1, x_2, a)$ at the $(n+1)^{\text{th}}$ iteration is given by

$$V_i^{(n+1)}(x_1, x_2, a) = (TV_i^{(n),\mu_{\bar{i}}^{(n+1)}})(x_1, x_2, a) - (TV_i^{(n),\mu_{\bar{i}}^{(n+1)}}(x_1', x_2', a')),\ i = 1,2, \text{ for all } (x_1, x_2, a) \tag{3.6}$$

If the above iteration converges to some value $V_i^*(x_1, x_2, a)$, for all states (x_1, x_2, a), then the $V_i^*(x_1, x_2, a)$ must satisfy

$$(TV_i^{\mu_{\bar{i}}^*}(x_1', x_2', a')) + V_i^*(x_1, x_2, a) = (TV_i^{\mu_{\bar{i}}^*})(x_1, x_2, a) \tag{3.7}$$

which from (3.3) implies that $J_i^* = (TV_i^{\mu_{\bar{i}}^*}(x_1', x_2', a'))$. The condition for convergence is

$$\left| \max_{(x_1, x_2, a)} \Delta V_i^{(n+1)}(x_1, x_2, a) - \min_{(x_1, x_2, a)} \Delta V_i^{(n+1)}(x_1, x_2, a) \right| < \varepsilon (TV_i^{(n),\mu_{\bar{i}}^{(n+1)}}(x_1', x_2', a')),\ i = 1,2 \tag{3.8}$$

where ε is a small chosen scalar and $\Delta V_i^{(n+1)}(x_1, x_2, a) \equiv V_i^{(n+1)}(x_1, x_2, a) - V_i^{(n)}(x_1, x_2, a)$.

A detailed version of the value iteration algorithm is shown in the Appendix of this chapter.

3.3 Numerical Results

To study the influence of the model parameters on the optimal ordering policies and long-run average profit of the suppliers, we implemented the value iteration method described in the previous section on several problem instances with 2 and 4 credibility states. In all instances, the selling price per item is the same for both suppliers, i.e. $r_1 = r_2 \equiv r$, since the suppliers do not compete on price. Moreover, in all instances, the customer's demand in each period, w, is equal to $\hat{w} - 1$, where $\hat{w}$ is geometrically distributed with mean $1/\rho$, i.e. w has the following probability distribution

$$\begin{gathered} p(w) = \rho(1-\rho)^w, \quad 0 < \rho < 1 \\ T(w) = 1-(1-\rho)^{w+1} \\ \theta = E[w] = \frac{1-\rho}{\rho} \end{gathered} \tag{3.9}$$

We also examined the same instances with Poisson distributed demand and found similar results.

Our most important finding is that in all instances the optimal ordering policy of both suppliers is an order-up-to policy, where the optimal order-up-to level of each supplier depends only on his credibility level and not on the inventory surplus/backorder level of the other supplier. In other words, all our numerical experiments showed that the optimal control law of supplier i is given by

$$\mu_i^*(x_1, x_2, a) = \begin{cases} 0, & x_i > s_i(a) \\ s_i(a) - x_i, & x_i \le s_i(a) \end{cases}$$

where $s_i(a)$ is the optimal order-up-to level of supplier i when the credibility level of supplier 1 is a.

Table 3-1 shows the parameter values and the corresponding performance measures, i.e. the optimal order-up-to levels and average profit, for both suppliers, for 19 problem instances with 2 credibility states, low and high, i.e. $a \in \{0,1\}$. The last two columns of Table 3-1 show the number of iterations (N) and clock time in seconds that it took for the value iteration to converge on an AMD Athlon 64 3000+ @1.8 GHz notebook, for $\varepsilon = 0.00001$.

The 19 instances in Table 3-1 are clustered into five groups: 1-3, 4-6, 7-10, 11-14, and 15-19.

Table 3-1. Input parameters and results for 19 instances with 2 credibility states

#	ρ	r	c_1	c_2	h_1	h_2	$q_1(0)$	$q_1(1)$	$s_1(0)$	$s_1(1)$	$s_2(0)$	$s_2(1)$	J_1^*	J_2^*	N	Comp. Time (sec)
1	0.35	10	5	5	0.01	0.01	0.4	0.6	8	8	8	8	4.57	4.57	80	680.90
2	0.35	10	5	7	0.01	0.2	0.4	0.6	7	8	1	0	5.1	2.4	91	690.00
3	0.35	10	5	7	0.2	0.01	0.4	0.6	1	2	7	6	4.09	2.90	79	689.90
4	0.35	15	5	5	0.01	0.01	0.4	0.6	9	10	10	9	9.19	9.19	76	701.57
5	0.35	25	5	5	0.01	0.01	0.4	0.6	11	11	11	11	18.47	18.47	70	721.48
6	0.35	35	5	5	0.01	0.01	0.4	0.6	12	12	12	12	27.74	27.74	67	679.82
7	0.7	10	5	5	0.01	0.01	0.4	0.6	2	2	2	2	1.05	1.05	250	2431.35
8	0.6	10	5	5	0.01	0.01	0.4	0.6	3	3	3	3	1.63	1.63	173	1497.17
9	0.5	10	5	5	0.01	0.01	0.4	0.6	4	4	4	4	2.46	2.46	126	949.26
10	0.3	10	5	5	0.01	0.01	0.4	0.6	9	10	10	9	5.74	5.74	69	723.96
11	0.35	10	5	5	0.01	0.01	0.0	0.2	0	0	9	9	0.12	9.08	3049	29610.32
12	0.35	10	5	5	0.01	0.01	0.1	0.3	4	5	9	9	1.11	8.04	347	3223.34
13	0.35	10	5	5	0.01	0.01	0.2	0.4	6	7	9	9	2.66	6.88	168	1564.15
14	0.35	10	5	5	0.01	0.01	0.3	0.5	7	7	9	8	3.41	5.73	109	1019.71
15	0.35	10	5	5	0.01	0.01	0.2	0.2	0	0	0	0	1.85	7.42	213	1916.25
16	0.35	10	5	5	0.01	0.01	0.2	0.4	6	7	9	9	2.66	6.88	168	1564.15
17	0.35	10	5	5	0.01	0.01	0.2	0.6	8	9	11	9	3.00	6.09	131	1241.79
18	0.35	10	5	5	0.01	0.01	0.2	0.8	10	12	12	10	4.53	4.53	97	984.64
19	0.35	10	5	5	0.01	0.01	0.2	1.0	10	10	7	0	8.92	0.21	1835	18200.90

In the first three groups of instances (1-3, 4-6, and 7-10), the probability of choosing either supplier when his credibility level is low is the same and equal to 0.4; therefore, in these instances, the customer exhibits a symmetric goodwill behavior towards the two suppliers.

More specifically, in the first group of instances (1-3), we study the effect of the cost parameters, $h_i, c_i,\ i = 1, 2$, on the suppliers' performance. In instance 1, both suppliers have the same ordering and inventory holding cost parameters; therefore their performance measures are identical. In instance 2, supplier 2's cost parameters are higher than those of supplier 1, which are kept at the same values as in instance 1. As a result, supplier 2's optimal order-up-to levels drop dramatically, while those of supplier 1 remain almost the same. Consequently, supplier 1 gains some market share from supplier 2 and thus increases his average profit. Supplier 2's profit, on the other hand, drops to almost half its value, because of his higher ordering and inventory holding costs as well as his loss of market share. In instance 3, supplier 2's ordering cost is higher than that of supplier 1, which is kept at the same value as in instance 1. At the same time, supplier 2's inventory holding cost remains at the same value as in instance 1, while that of supplier 1 is increased. As a result, supplier 1's optimal order-up-to levels drop dramatically, while those of supplier 2 drop only slightly. Consequently, supplier 2 gains some market share from supplier 1 but his average profit still drops quite sharply, because his gross profit margin (selling price – procurement cost) is reduced. Supplier 1's profit also drops, because of his higher inventory holding costs as well as his loss of market share.

In the second and third group of instances (4-6 and 7-10), we study the effect of the selling price *r* and the demand distribution parameter *ρ*, respectively, on the suppliers' performance. In all these instances, both suppliers have the same cost parameters; therefore their performance measures are identical. We see that as the selling price *r* or demand distribution parameter *ρ* increases, both suppliers increase their order-up-to levels, so as not to loose market share, and still gain higher profits.

Finally, in the fourth and fifth groups of instances, (11-14 and 15-19), both suppliers have the same cost parameters, but the probability of choosing a supplier when his credibility level is low, is no longer the same for both suppliers, as was the case in the first three groups of instances; therefore, in these instances, the customer exhibits an asymmetric goodwill behavior towards the two suppliers.

More specifically, in the fourth group of instances (11-14), we study the effect of increasing the probabilities $q_1(0)$ and $q_1(1)$ while keeping their difference constant and equal to 0.2. As $q_1(0)$ and therefore $q_1(1)$ increases, supplier 1's market share – and therefore his gross profit – increases. This allows him to raise his optimal order-up-to levels, in order to increase the long-run probability that his credibility level is high. Supplier 2 behaves exactly the opposite way. In all these instances, the customer's goodwill is biased toward supplier 2, because the probability that he chooses supplier 2, when supplier 2's credibility level is low, is higher that the probability that he chooses supplier 1, when supplier 1's credibility level is low. For this reason, supplier 2's order-up-to levels and the resulting average profit are higher than those of supplier 1. Note that in instance 11, $s_1(0)=0$. This is absolutely reasonable, because the probability that the customer will choose supplier 1 when his credibility level is low is zero, since $q_1(0)=0$; therefore, it makes no sense for supplier 1 to hold any inventory when his credibility level is low.

Finally, in the fifth group of instances (15-19), we study the effect of increasing probability $q_1(1)$ while keeping $q_1(0)$ constant and equal to 0.2. We see that when $q_1(1)=q_1(0)$, both suppliers have zero order-up-to levels, i.e. they never hold any inventory. This is absolutely reasonable, because when $q_1(1)=q_1(0)$, neither supplier ever gains or looses customer goodwill (in the sense of changing the probability of being chosen in the future) after providing a good or poor service, respectively. With this in mind and given that a stockout incur no direct penalty cost, there is no motive for either supplier to hold inventory. As $q_1(1)$ increases, however, the optimal order-up-to level of both suppliers increases. In instances 15-17, the customer's goodwill is biased toward supplier 2. For this reason, supplier 2's order-up-to levels and the resulting average profit are higher than those of supplier 1. In instance 18, which is in fact similar to instance 1, the customer's goodwill behavior toward both suppliers is completely symmetric; hence the performance measures of both suppliers are the same. However, the order-up-to levels of both suppliers are higher than those in instance 1. This is because the gain and loss of customer goodwill (in the sense of changing the probability of being chosen in the future) after providing a good or poor service, respectively, is bigger in instance 18 than in instance 1.

A general remark is that in all instances in Table 3-1, the optimal order-up-to level of each supplier when his credibility level is high is no less than his optimal order-up-to level when his credibility level is low, i.e. $s_1(1) \geq s_1(0)$ and $s_2(0) \geq s_2(1)$. This implies that a supplier must hold at least as many items in inventory when his credibility is high than when it is low, which further suggests that more money tied in inventories is needed to *keep* a position of high credibility, where the expected rewards are high, than to *gain* such a position. The above observation is no longer true when more than two credibility states, as we will see next.

We also run 3 problem instances with 4 credibility states each, i.e. $a \in \{0,1,2,3\}$, in which we study the effect of the shape of $q_1(a)$ on the suppliers' performance measures. In all instances the cost parameters of both suppliers are the same, i.e. $c_1 = c_2 \equiv c$ and $h_1 = h_2 \equiv h$. Also, in all instances $q_1(a) = \overline{q}_1(3-a) \equiv q_2(3-a)$, $a = 0,1,2,3$; therefore, the customer's goodwill behavior towards the two suppliers is completely symmetric. Table 3-2 shows the parameter values and the corresponding performance measures for supplier 1; those of supplier 2 are completely symmetric, i.e. $s_2(a) = s_1(3-a)$, $a = 0,1,2,3$, and $J_2^* = J_1^*$, so they are not shown.

Table 3-2. Input parameters and results for three instances with 4 credibility states

#	ρ	r	c	h	$q_1(0)$	$q_1(1)$	$q_1(2)$	$q_1(3)$	$s_1(0)$	$s_1(1)$	$s_1(2)$	$s_1(3)$	J_1^*	N	Comp. time (sec)
1	0.35	10	5	0.01	0.2	0.4	0.6	0.8	10	12	13	13	4.52	125	2566.82
2	0.35	10	5	0.01	0.2	0.3	0.7	0.8	10	12	14	13	4.52	139	2849.93
3	0.35	10	5	0.01	0.2	0.2	0.8	0.8	10	12	15	13	4.51	190	3924.43

In all three instances in Table 3-2, the lowest and highest values of $q_1(a)$ are 0.2 and 0.8, respectively. The three instances differ in the shape of $q_1(a)$. While in all three instances, $q_1(a)$ is in a sense “S”-shaped, it is flatter in instance 2 than in instance 3 (where it is piece-wise linear) and completely flat (linear) in instance 1. For this reason, there is a minor difference in the optimal performance measures between the three instances. This difference is that the optimal order-up-to level corresponding to a = 3 is slightly higher in instance 2 than in instance 1 and even higher in instance 3. A closer observation of the results, reveals that in instances 2 and 3, $s_1(a)$ is no longer non-decreasing in a, as was the case in all problem instances with two credibility states, shown in Table 3-1. Instead, $s_1(a)$ is increasing in a for $a \le 2$, but it decreases for $a \ge 2$. This can be explained by the fact that, in these two instances, the loss in the customer’s goodwill towards supplier 1 is higher if supplier 1’s credibility level drops from 2 to 1 than if drops from 3 to 2.

3.4 The Case of Two Extreme Credibility Levels

Thus far in this chapter we have assumed that an equilibrium solution the two-supplier game exists, but we have not proved that it does. This is in fact quite a formidable task. To shed some light into the question of existence and uniqueness of an equilibrium, in this section we will restrict our attention to the special case where each supplier has only two extreme credibility levels, a low and a high, such that when in the low level, he is never chosen by the customer, and when in the high level, he is always chosen by the customer. In other words, $a^t \in \{0,1\}$ and $q_1(0) = 0$, $q_1(1) = 1$. In addition, we assume that each supplier uses a credibility level-dependent order-up-to policy. Our goal is to find the optimal order-up-to levels, $s_i(0)$ and $s_i(1)$,of each supplier i, i = 1,2, at equilibrium.

Given that when the credibility level of supplier i is low, the customer will definitely not chose him, it is obvious that there is no benefit for supplier i to keep any inventory when his credibility state is low; therefore his order-up-to level should be zero, i.e., $s_1^*(0) = s_2^*(1) = 0$.

Suppose then that $s_1(0) = s_2(1) = 0$ and $s_1(1) = s_1$, $s_2(0) = s_2$, for some non-negative numbers s_1, s_2. Figure 3-1 shows a sample path of the inventory surplus/backlog levels of both suppliers. At the beginning of period 0, the credibility level of supplier 1 is 1, i.e. $a^0 = 1$; therefore supplier 1, knowing that the customer will chose him, orders up to s_1. Supplier 2, on the other hand, knowing that the customer will not choose him, orders up to zero. The customer indeed chooses supplier 1 and demands from him a number of items. Supplier 1 satisfies all the demand from inventory and so his inventory level at the beginning of period 1 drops to x_1^1, where $x_1^1 \geq 0$, so it is a surplus. The inventory level of supplier 2, on the other hand, remains at zero, i.e. $x_2^1 = 0$. Since supplier 1 satisfied all the demand, his credibility level remains at 1 at the beginning of period 1, i.e. $a^1 = 1$; therefore, once again, supplier 1 orders up to s_1. Supplier 2, on the other hand, orders up to zero, i.e. he orders no items at all. The same exact sequence of events holds for the next period. Therefore, at the beginning of period 2, right after ordering, the inventory level of supplier 1 is at s_1, while that of supplier 2 is at zero. Also $a^2 = 1$. The customer once again chooses supplier 1, but this time supplier 1 is unable to meet the demand. His inventory level drops to x_1^3, where $x_1^3 < 0$, so it is a backlog. Since supplier 1 did not satisfied all the demand, his credibility level drops to 0 at time 3, i.e. $a^3 = 0$; therefore, this time, knowing that the customer will not chose him in that period, he orders up to zero. Supplier 2, on the other hand, knowing that the customer will chose him in that period, orders up to s_2. Indeed, the customer chooses supplier 2 in period 3 and in the following periods, until supplier 2 fails to satisfy all the demand in period 5. So, in period 6, the customer switches back to supplier 1 and the cycle is repeated.

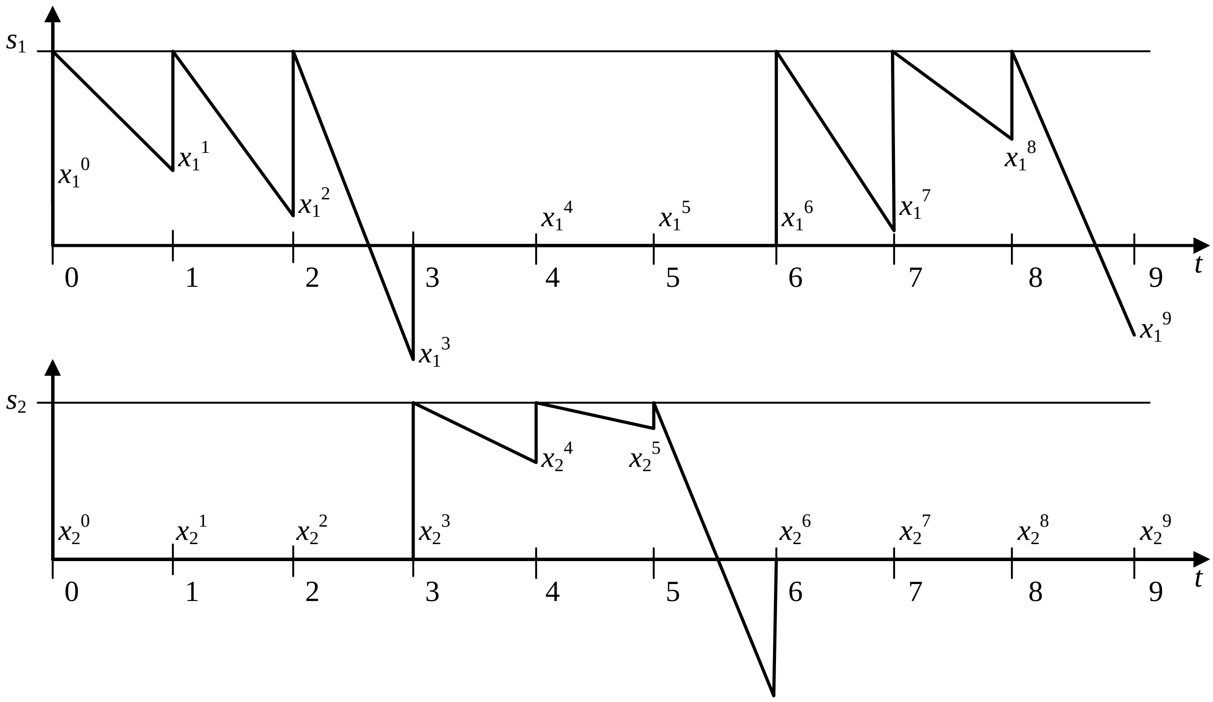

Figure 3-1. Sample path of both suppliers’ inventory surplus/backlog

If we look at the evolution of the inventory surplus/backlog of supplier 1, we will notice that he receives no rewards and incurs no costs during the periods in which his credibility level is low, i.e., periods 3-5 (the same holds for supplier 2). If we cut these periods off the graph and paste together the ends of the rest of the sample path, we will end up with the sample path shown in Figure 3-2. This is the sample path of a standard single-newsvendor model, where the newsvendor uses an order-up-to s_1 policy and earns a profit of $(r_1 - c_1)w - h_1(s_1 - w)^+$ per period. In other words, supplier 1 earns a profit of $(r_1 - c_1)w - h_1(s_1 - w)^+$ during the periods where $a = 1$, and earns no profit at all during the periods where $a = 0$. Therefore, his long-run average profit is equal to the expected value of $(r_1 - c_1)w - h_1(s_1 - w)^+$ multiplied by steady-state probability that $a^t = 1$.

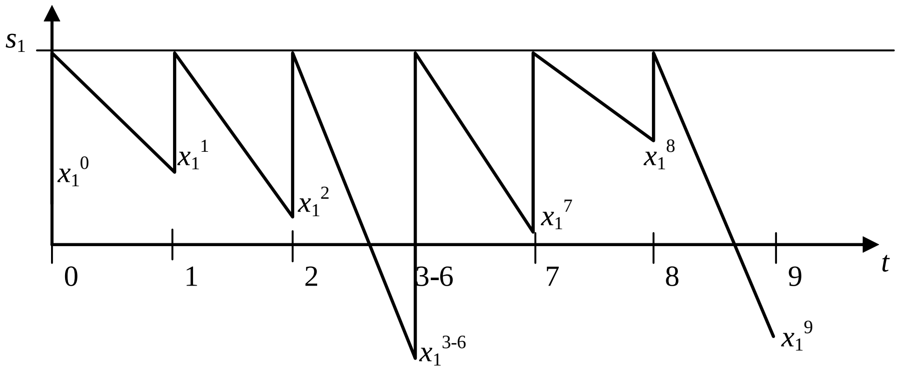

Figure 3-2. Sample path of supplier 1's inventory surplus/backlog

It is easy to see that $\{a^t,\ t = 0,1,\ldots\}$ is a simple two-state Markov chain with one-step transition probabilities

$$P_{00} = T(s_2),\ P_{01} = \bar{T}(s_2)$$
$$P_{10} = \bar{T}(s_1),\ P_{11} = T(s_1)$$

The steady-state probabilities of this Markov chain are given by

$$\pi_0 = \frac{\bar{T}(s_1)}{\bar{T}(s_1) + \bar{T}(s_2)},\ \pi_1 = \frac{\bar{T}(s_2)}{\bar{T}(s_1) + \bar{T}(s_2)} \tag{3.10}$$

Let $J_i(s_1, s_2)$ be the long-run average profit of supplier *i*, given that $s_1(0) = s_2(1) = 0$ and $s_1(1) = s_1,\ s_2(0) = s_2$. Then, based on our discussion above, $J_i(s_1, s_2),\ i = 1,2$ is given by

$$J_1(s_1, s_2) = \frac{\bar{T}(s_2)}{\bar{T}(s_1) + \bar{T}(s_2)}\Big((r_1 - c_1)\theta - h_1 E\big[(s_1 - w)^+\big]\Big) \tag{3.11}$$

$$J_2(s_1,s_2)=\frac{\bar{T}(s_1)}{\bar{T}(s_1)+\bar{T}(s_2)}\Big((r_2-c_2)\theta-h_2E\big[(s_2-w)^+\big]\Big) \tag{3.12}$$

Expressions (3.11) and (3.12) are typical payoff functions in a two-player competitive game. It can be easily verified that

$$J_1(0,s_2)=\frac{\bar{T}(s_2)}{1+\bar{T}(s_2)}(r_1-c_1)\theta \tag{3.13}$$

$$\lim_{s_1\to\infty} J_1(s_1,s_2)=-\infty \tag{3.14}$$

Expression (3.13) implies that supplier 1's average profit is positive when $s_1=0$, as long as $s_2<\infty$, and expression (3.14) states that supplier 1's profit tends to $-\infty$ as s_1 tends to infinity. Similar expressions can be written for $J_2(s_1,0)$ and $\lim_{s_2\to\infty} J_2(s_1,s_2)$.

The goal of supplier *i* is to set his order-up-to level s_i so as to maximize his payoff function, given that his competitor uses order-up-to level $s_{\bar{i}}$. Let $s_i^*(s_{\bar{i}})$ be the optimal order-up-to level of supplier *i*, given that his competitor uses order-up-to level $s_{\bar{i}}$, i.e.,

$$s_i^*(s_{\bar{i}})\equiv \arg\max_{s_i} J_i(s_1,s_2) \tag{3.15}$$

We will refer to $s_i^*(s_{\bar{i}})$ as supplier *i*'s "best response function" to his competitor's decision variable $s_{\bar{i}}$ (see, Cahon and Netessine, 2005). A question of theoretical and practical interest is, does a Nash equilibrium (NE) exist, i.e. is there a pair (s_1^*,s_2^*) such that s_2^* is a best response to s_1^* and vice versa? According to Theorem 1 in Cahon and Netessine (2005), if the decision space of each player is compact (closed and bounded) and the payoff function is continuous and quasi-concave, then there exists at least one pure NE.

In our model, the decision space of each supplier is not bounded from above; however we could easily bound it with some large enough finite number to represent the upper bound on the demand distribution. That bound would not affect any of the order quantities and so the transformed game would behave just like the original game with an unbounded decision space. What is difficult to show is that payoff function of each player is quasi-concave. To do this, we would have to prove that the second partial derivative of each supplier's payoff function with respect to his order-up-to level is non-positive. To see why this is a formidable task, consider the first partial derivative of any one supplier, say supplier 1, with respect to his order-up-to level. This is given by

$$\frac{\partial J_1(s_1,s_2)}{\partial s_1}=\frac{f(s_1)\bar{T}(s_2)}{\left(\bar{T}(s_1)+\bar{T}(s_2)\right)^2}\Big((r_1-c_1)\theta-h_1E\left[(s_1-w)^+\right]\Big)-h_1\frac{T(s_1)\bar{T}(s_2)}{\bar{T}(s_1)+\bar{T}(s_2)} \tag{3.16}$$

It can be easily verified that

$$\left.\frac{\partial J_1(s_1,s_2)}{\partial s_1}\right|_{s_1=0}=\frac{f(0)\bar{T}(s_2)}{\left(1+\bar{T}(s_2)\right)^2}(r_1-c_1)\theta \tag{3.17}$$

$$\lim_{s_1\to\infty}\frac{\partial J_1(s_1,s_2)}{\partial s_1}=-h_1 \tag{3.18}$$

Similar expressions can be derived for supplier 2.

Expression (3.17) implies that the derivative of the average profit of supplier 1 is positive at $s_1=0$, as long as $s_2<\infty$. This further implies that $s_1^*(s_2)>0$. Expression (3.18) states that the derivative of the average profit of supplier 1 with respect to s_1 is negative as s_1 tends to infinity. This means that $J_1(s_1,s_2)$ is increasing in s_1 at $s_1=0$ and decreasing as s_1 tends to infinity; however, we do not know if $J_1(s_1,s_2)$ is unimodal. To find out, we would have to determine the sign of the second partial derivative of $J_1(s_1,s_2)$ with respect to s_1; however, no firm conclusions can be reached about the sign of $\partial J_1^2(s_1,s_2)/\partial s_1\partial s_1$, whose expression is too complicated to even display it here.

Another option to see whether there exists al least on NE would be to investigate whether the game in our model is supermodular, in view of Theorem 3 in Cachon and Netessine (2005), which states that in a supermodular game[1] there exists at least one NE. The second partial cross derivative of $J_1(s_1,s_2)$ is given by the following expression:

$$\begin{aligned}\frac{\partial J_1^2(s_1,s_2)}{\partial s_1\partial s_2}=&\frac{f(s_2)}{\left(\bar{T}(s_1)+\bar{T}(s_2)\right)^3}\Big\{\big(T(s_1)-T(s_2)\big)\Big((r_1-c_1)\theta-h_1E\left[(s_1-w)^+\right]\Big)\\&-h_1T(s_1)\Big(T(s_1)-T(s_2)+T(s_1)\big(\bar{T}(s_1)+\bar{T}(s_2)\big)\Big)\Big\}\end{aligned}$$

Unfortunately, no firm conclusions can be reached about the sign of the above expression.

It appears then that there is no easy way to obtain any general results concerning the existence of a NE. For this reason, we will next limit our attention to the special case where the demand distribution is discrete and is given by (3.9).

[1] A supermodular game is a game in which all the players' payoff functions are supermodular. Supplier 1's payoff function $J_1(s_1,s_2)$ is supermodular if $\partial J_1^2(s_1,s_2)/\partial s_1\partial s_2\geq 0$, for all s_1,s_2.

Suppose then that the distribution of the customer's demand is given by (3.9). It is easy to verify that

$$\begin{aligned} E\left[(s_1-w)^+\right] &= s_1 - \frac{1-\rho}{\rho}\left(1-(1-\rho)^{s_1}\right) \\ &= s_1 - \theta T(s_1-1) \end{aligned} \tag{3.19}$$

After substituting $\bar{F}(\cdot)$ from (3.9) and $E[(s_1-w)^+]$ from (3.19) into the payoff of supplier 1, given by (3.11), we obtain

$$\begin{aligned} J_1(s_1,s_2) &= \frac{\bar{T}(s_2)}{\bar{T}(s_1)+\bar{T}(s_2)}\left((r_1-c_1)\theta - h_1\left(s_1-\theta T(s_1-1)\right)\right) \\ &= \frac{(1-\rho)^{s_2}}{\rho\left((1-\rho)^{s_1}+(1-\rho)^{s_2}\right)}\left((r_1-c_1)(1-\rho) - h_1\left(\rho s_1 - (1-\rho)(1-(1-\rho)^{s_1})\right)\right) \end{aligned} \tag{3.20}$$

The order-up-to levels of the two suppliers, s_1, s_2, are discrete numbers. To examine the shape of $J_1(s_1,s_2)$ with respect to s_1 we will examine the difference in supplier 1's payoff function for two adjacent order-up-to levels, s_1-1 and s_1. After some algebraic manipulations this difference can be written as follows

$$\begin{aligned} &J_1(s_1,s_2) - J_1(s_1-1,s_2) = \\ &\quad -\frac{h_1\left[s_1\rho(1-\rho)^{s_1} + (1-\rho)^{s_2+1}(1-(1-\rho)^{s_1})\right] - (r_1-c_1)(1-\rho)^{s_1+1}}{(1-\rho)^{-s_2}\left((1-\rho)^{s_1}+(1-\rho)^{s_2}\right)\left((1-\rho)^{s_1}+(1-\rho)^{s_2+1}\right)} \end{aligned} \tag{3.21}$$

If $J_1(s_1,s_2)$ is concave in s_1, then the best response function of supplier 1 is given by

$$s_1^*(s_2) = \arg\min_{s_1:s_1 \text{ int}}\left\{J_1(s_1,s_2) - J_1(s_1-1,s_2) \le 0\right\} - 1$$

In order for the difference given by (3.21) to be non-positive, the term inside the brackets in the numerator must be non-negative, because the denominator is positive. In other words,

$$s_1^*(s_2) = \arg\min_{s_1:s_1 \text{ int}}\left\{h_1\left[s_1\rho(1-\rho)^{s_1} + (1-\rho)^{s_2+1}(1-(1-\rho)^{s_1})\right] - (r_1-c_1)(1-\rho)^{s_1+1} \ge 0\right\} - 1$$

If we divide both sides of the condition inside the brackets in the above expression by $h_1(1-\rho)^{s_1+1}$, which is positive, and perform some further algebraic manipulations, we obtain the following simpler expression

$$s_1^*(s_2) = \arg\min_{s_1:s_1 \text{ int}}\left\{\frac{\rho}{1-\rho}s_1 + (1-\rho)^{s_2-s_1} \ge (1-\rho)^{s_2} + \frac{r_1-c_1}{h_1}\right\} - 1 \tag{3.22}$$

A similar expression can be derived for $s_2^*(s_1)$.

Note that the left-hand side of the condition inside the brackets in expression (3.22) is equal to $(1-\rho)^{s_2}$ when $s_1 = 0$, so it is clearly smaller than the right-hand side. As s_1 increases, however, the left-hand side increases monotonically – in fact with an increasing rate – so the above expression has exactly one solution, which also proves that $J_1(s_1, s_2)$ is concave in s_1; therefore according to Theorem 1 in Cachon and Netessine (2005), there exists at least one pure NE. Next, we will prove that if s_1 and s_2 are continuous, there exists a *unique* NE.

Suppose we relax the integer constraints on s_1 and s_2, and allow s_1 and s_2 to be continuous. Then we can draw the left-hand side (lhs) and the right-hand side (rhs) of the condition inside the brackets in expression (3.22) as continuous functions of s_1, for different values of s_2. Figure 3-3 shows these functions for four different values of s_2, namely 0, s_2', $s_2' + \Delta s_2'$, and ∞, where $0 < s_2' < \infty$ and $\Delta s_2' > 0$. The rhs's are represented by the horizontal dotted lines, and the lhs's are shown by the solid increasing curves. The point where the lhs intersects – i.e. is equal to – the rhs, for a given s_2, is $s_1^*(s_2)$. On the *x*-axis of the graph, we show $s_1^*(s_2')$ – which we denote by s_1' – and $s_1^*(s_2' + \Delta s_2')$. It is clear from the graph that the mapping $s_1^*(s_2)$ is increasing in s_2; therefore $s_1^*(s_2' + \Delta s_2') \equiv s_1^*(s_2') + \Delta s_1' = s_1' + \Delta s_1'$, for some positive $\Delta s_1'$.

According to Theorem 4 in Cachon and Netessine (2005), if the best response function $s_1^*(s_2)$ is a contraction on the entire s_2 space (and similarly $s_2^*(s_1)$ is a contraction on the entire s_1 space), then there exists a unique NE. In our case, to show that $s_1^*(s_2)$ is a contraction on the entire s_2 space, we have to show that $\Delta s_1' < \Delta s_2'$.

As we mentioned earlier, s_1' and $s_1' + \Delta s_1'$ are the points that satisfy the condition "lhs = rhs" for $s_2 = s_2'$ and $s_2 = s_2' + \Delta s_2'$, respectively, i.e. s_1' and $s_1' + \Delta s_1'$ satisfy

$$\frac{\rho}{1-\rho} s_1' + (1-\rho)^{s_2'-s_1'} = (1-\rho)^{s_2'} + \frac{r_1 - c_1}{h_1}$$

$$\frac{\rho}{1-\rho}(s_1' + \Delta s_1') + (1-\rho)^{s_2'-s_1'+(\Delta s_2'-\Delta s_1')} = (1-\rho)^{s_2'+\Delta s_2'} + \frac{r_1 - c_1}{h_1}$$

If we subtract the first from the second equation above, we have

$$\frac{\rho}{1-\rho}\Delta s_1' + (1-\rho)^{s_2'-s_1'+(\Delta s_2'-\Delta s_1')} - (1-\rho)^{s_2'-s_1'} = (1-\rho)^{s_2'+\Delta s_2'} - (1-\rho)^{s_2'}$$

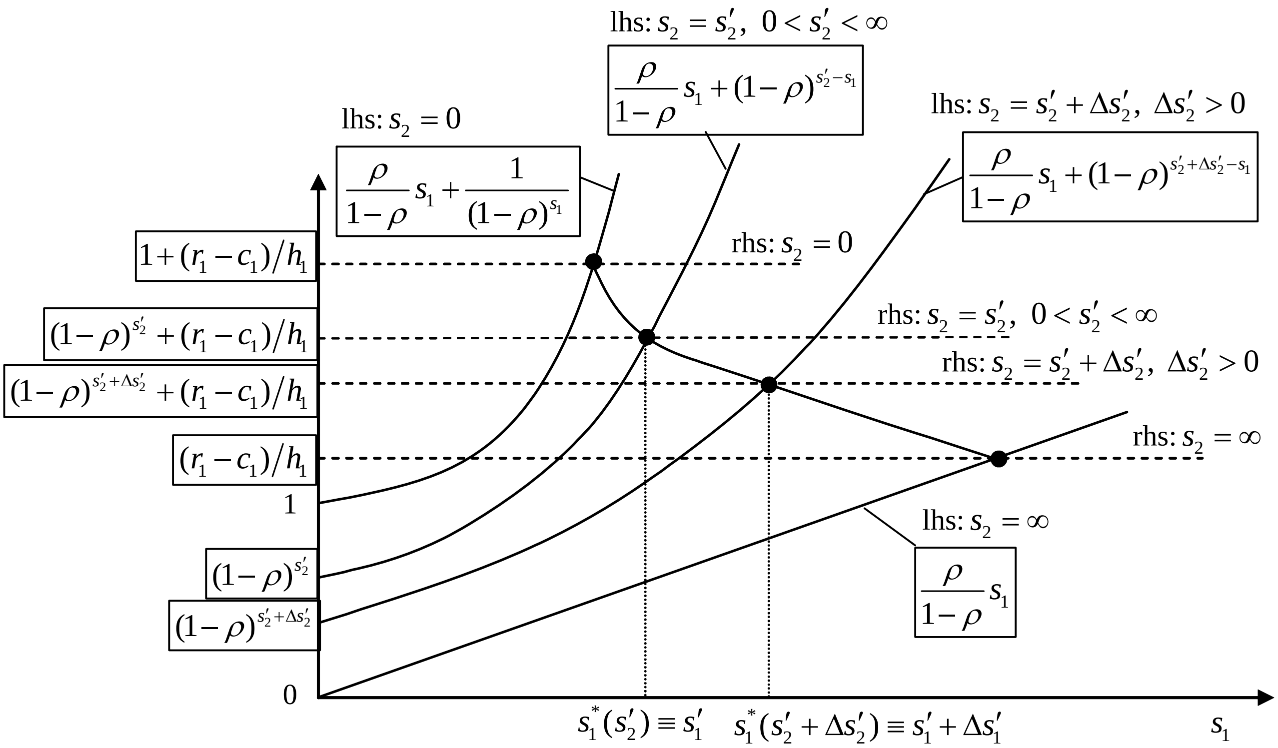

Figure 3-3. Left-hand side and right-hand side of the condition in (3.22) as functions of s_1, for different values of s_2

It is clear that the rhs of the above equality is negative and that the first term of the lhs is positive; therefore, in order for the equality to hold, the term $(1-\rho)^{s_2'-s_1'+(\Delta s_2'-\Delta s_1')} - (1-\rho)^{s_2'-s_1'}$ must be negative. But in order for this to happen, $\Delta s_2' - \Delta s_1'$ must be positive, which means that $\Delta s_1' < \Delta s_2'$. Consequently, $s_1^*(s_2)$ is a contraction on the entire s_2 space.

The same exact arguments can be used to show that $s_2^*(s_1)$ is also a contraction on the entire s_1 space. This guarantees that there exits a unique NE.

To study the influence of the model parameters on the optimal performance of the two suppliers, we found the optimal order-up-to levels at equilibrium and the corresponding optimal average profits of the two suppliers for 10 problem instances. The model parameters that we used for these instances are identical to those used in instances 1-10 in Table 3-1, except for $q_1(0)$ and $q_1(1)$, which are equal to 0 and 1, respectively. The input parameters and the results for the 10 instances are shown in Table 3-3. The optimal order-up-to levels at equilibrium, s_1^{NE}, s_2^{NE}, were found by performing a fixed point iteration to simultaneously solve

$$s_1^{NE} = s_1^*(s_2^{NE})$$
$$s_2^{NE} = s_2^*(s_1^{NE})$$

where $s_1^*(\cdot)$ is given by (3.22) and $s_2^*(\cdot)$ by a symmetric expression. J_1^* is then equal to $J_1(s_1^{NE}, s_2^{NE})$, where $J_1(\cdot,\cdot)$ is given by (3.20), and J_2^* is equal to $J_2(s_1^{NE}, s_2^{NE})$, where $J_2(\cdot,\cdot)$ is given by a symmetric expression. Although we proved that if s_1 and s_2 are continuous there exists a unique NE, in all instances that we examined we found exactly two NE points, namely (s_1^{NE}, s_2^{NE}) and $(s_1^{NE}+1, s_2^{NE}+1)$. This is because s_1 and s_2 were constrained to be integers. In Table 3-3, we chose to display the smallest NE, namely (s_1^{NE}, s_2^{NE}), because this yields a higher average profit for both suppliers.

Table 3-3. Input parameters and results for 10 instances with 2 extreme credibility states

#	ρ	r	c_1	c_2	h_1	h_2	s_1^{NE}	s_2^{NE}	J_1^*	J_2^*
1	0.35	10	5	5	0.01	0.01	925	925	0.0271	0.0271
2	0.35	10	5	7	0.01	0.2	41	27	8.8730	0.0013
3	0.35	10	5	7	0.2	0.01	46	58	0.0026	4.9817
4	0.35	15	5	5	0.01	0.01	1854	1854	0.0250	0.0250
5	0.35	25	5	5	0.01	0.01	3711	3711	0.0257	0.0257
6	0.35	35	5	5	0.01	0.01	5568	5568	0.0264	0.0264
7	0.7	10	5	5	0.01	0.01	212	212	0.0136	0.0136
8	0.6	10	5	5	0.01	0.01	331	331	0.0150	0.0150
9	0.5	10	5	5	0.01	0.01	497	497	0.0200	0.0200
10	0.3	10	5	5	0.01	0.01	1163	1163	0.0300	0.0300

From the results in Table 3-3, we see that in all symmetric instances, namely instances 1 and 4-10, the optimal order-up-to levels of both suppliers are extremely large, which suppresses their average profit slightly above zero. Only in instances 2 and 3 one of the two suppliers (the one with the lower inventory holding cost) achieves an average profit which is not close to zero.

3.5 Conclusions

We analyzed a discrete-time infinite horizon inventory model in which two suppliers compete for a single customer on product availability. We formulated the problem of finding optimal stationary ordering policies for both suppliers at equilibrium as a stochastic dynamic game, and we numerically solved the resulting optimality conditions for several instances of this problem. The results indicate that both suppliers must follow the same type of policy, which can be characterized as an order-up-to policy. The order-up-to levels are generally different for each supplier and as the numerical examples suggested, they are independent of the competitor's inventory position.

It may not be that difficult to prove that the optimal ordering policy for each supplier, given that his competitor uses a stationary ordering policy, is an order-up-to policy. What is certainly much more demanding to prove is the existence and uniqueness of a NE, as the analysis of the simple two-credibility-level system in Section 3.4 showed.

The proposed inventory model could be extended and modified in a number of ways. Firstly, we could consider the situation where there is no information sharing between the two suppliers. In this case each supplier must decide his replenishment orders without knowing the inventory or service level of his competitor. Another meaningful extension of our model would be to set the selling values as decision variables. In this way the customer's next supplier choice would be affected not only by previous service levels, but by pricing policies as well. Finally, we could generalize the model for more than two suppliers.

Appendix

The complete value iteration algorithm outlined in Section 3.2 is as follows:

Main Routine:

```
int i, j;
#include "Implementation Routine.c"
int a = 0;    /* Credibility state of supplier 1 */
int X1 = 0;  /* Inventory level of supplier 1 */
int X2 = 0;  /* Inventory level of supplier 2 */
int g = 0;
int k = 0;

/* Main Routine */
main()
{FILE *fptr;
fptr = fopen("results.txt", "w");

/* Variable initializations */
V1diffmax = –1;
V2diffmax = –1;
V1diffmin = 1;
V2diffmin = 1;

/* Probabilities q1(α), α = 0,.., 3 */
```

```
P[0] = 0.2;
P[1] = 0.4;
P[2] = 0.5;
P[3] = 0.8;

/* Check for convergence according to (3.8) */
while (fabs(V1_{diffmax} – V1_{diffmin}) > fabs(e*T_{Table1}[X_1´ – X_{1min}][ X_2´ – X_{2min}][ a_1´]) || fabs(V2_{diffmax} –
V2_{diffmin}) > fabs(e*T_{Table2}[X_1´ – X_{1min}][ X_2´ – X_{2min}][ a_2´])){

    /* Run main value iteration step. * /
    for (a = 0; a <= M; a++){
        for (X_2 = X_{2min}; X_2 <= X_{2max}; X_2++){
            for (X_1 = X_{1min}; X_1 <= X_{1max}; X_1++){
                /* Set V1_{old}, V2_{old} equal to the differential profit function values from previous step. */
                V1_{old}[X_1 – X_{1min}][X_2 – X_{2min}][a] = V1[X_1 – X_{1min}][ X_2 – X_{2min}][a];
                V2_{old}[X_1 – X_{1min}][X_2 – X_{2min}][a] = V2[X_1 – X_{1min}][ X_2 – X_{2min}][a];
                /* Find new optimal control and differential profit mapping values from (3.5) */
                optU1U2(X_1, X_2, a); }}}

    for (a = 0; a <= M; a++){
        for (X_2 = X_{2min}; X_2 <= X_{2max}; X_2++){
            for (X_1 = X_{1min}; X_1 <= X_{1max}; X_1++){
                /* Find new differential profit function values from (3.6) */
                V1[X_1 – X_{1min}][X_2 – X_{2min}][a] = T_{Table1}[X_1 – X_{1min}][X_2 – X_{2min}][a] – T_{Table1}[X_1´ – X_{1min}][
                    X_2´ – X_{2min}][ a_1´];
                V2[X_1 – X_{1min}][X_2 – X_{2min}][a] = T_{Table2}[X_1 – X_{1min}][X_2 – X_{2min}][a] – T_{Table2}[X_1´ – X_{1min}][
                    X_2´ – X_{2min}][ a_2´];
                V1_{diffmax} = max(V1_{diffmax}, (V1[X_1 – X_{1min}][X_2 – X_{2min}][a] – V1_{old}[X_1 – X_{1min}][X_2 –
                    X_{2min}][a]));
                V2_{diffmax} = max(V2_{diffmax}, (V2[X_1 – X_{1min}][X_2 – X_{2min}][a] – V2_{old}[X_1 – X_{1min}][X_2 –
                    X_{2min}][a]));
                V1_{diffmin} = min(V1_{diffmin}, (V1[X_1 – X_{1min}][X_2 – X_{2min}][a] – V1_{old}[X_1 – X_{1min}][X_2 –
                    X_{2min}][a]));
                V2_{diffmin} = min(V2_{diffmin}, (V2[X_1 – X_{1min}][X_2 – X_{2min}][a] – V2_{old}[X_1 – X_{1min}][X_2 –
                    X_{2min}][a])); }}}

    printf("V1_{diffmax} – V1_{diffmin} = %Lf\t", V1_{diffmax} – V1_{diffmin});
    printf("e*T_{Table1'} = %Lf\n", e*T_{Table1}[X_1´ – X_{1min}][ X_2´ – X_{2min}][ a_1´]);
    printf("V2_{diffmax} – V2_{diffmin} = %Lf\t", V2_{diffmax} – V2_{diffmin});
    printf("e*T_{Table2'} = %Lf\n", e*T_{Table2}[X_1´ – X_{1min}][ X_2´ – X_{2min}][ a_2´]);}

print_UT_{Tables}();
fclose (fptr); }
```

Implementation Routine:

```
#include "Declaration Routine.h"
```

/ Routine that finds new $\mu_1(x_1, x_2, a)$ values from the 1st expression of (3.4), given old $\mu_2(x_1, x_2, a)$ values */*

```
void optU1(int X1, int X2, int a, int U2)
{long double temp1;
long double cumprob1;
long double prob1;
long double temp11;

T1temp = –10000;
for (U1 = – min(0, X1); U1 <= min(U1max, X1max – X1); U1++){
    Wmax1 = X1 + U1 – X1min;
    Wmax2 = X2 + U2 – X2min;
    temp1 = (long double) – C1*U1;

    if (P[a] != 0){
        cumprob1 = 0.0;
        temp11 = 0.0;
        for (w = 0; w <= X1 + U1; w++){
            prob1 = p*(pow((1 – p), w));
            cumprob1 = cumprob1 + prob1;
            temp11 = temp11 + (prob1)*(r1*w – h1*(X1 + U1 – w) + V1[(X1 + U1 – w) – X1min][ X2 +
              U2 – X2min][min((a + 1), M)]);}
        for (w = min(X1 + U1 + 1, Wmax1); w <= Wmax1; w++){
            prob1 = p*(pow((1 – p), w));
            cumprob1 = cumprob1 + prob1;
            temp11 = temp11 + (prob1)*(r1*w + V1[(X1 + U1 – w) – X1min][ X2 + U2 – X2min][max((a –
                1), 0)]);}

        temp11 = temp11 + (1 – cumprob1)*(r1*(X1 + U1 – X1min) + V1[(X1min) – X1min][ X2 + U2 –
            X2min][max((a – 1), 0)]);
        temp1 = temp1 + P[a]*temp11;}

    if (P[a] != 1){
        cumprob1 = 0.0;
        temp11 = 0.0;
        for (w = 0; w <= (X2 + U2); w++){
            prob1 = p*(pow((1 – p), w));
            cumprob1 = cumprob1 + prob1;
            temp11 = temp11 + (prob1)*(V1[(X1 + U1) – X1min][( X2 + U2 – w) – X2min][max((a – 1),
                0)] – h1*max(0, ( X1 + U1)));}
        for (w = min(X2 + U2 + 1, Wmax2); w <= Wmax2; w++){
            prob1 = p*(pow((1 – p), w));
            cumprob1 = cumprob1 + prob1;
            temp11 = temp11 + (prob1)*(V1[(X1 + U1) – X1min][( X2 + U2 – w) – X2min][min((a + 1),
                M)] – h1*max(0, ( X1 + U1)));}

        temp11 = temp11 + (1 – cumprob1)*(V1[(x1 + u1) – x1min][(x2min) – x2min][min((a + 1),
            M)] – h1*max(0, (x1 + u1)));
```

```
        temp1 = temp1 + (1 – P[a])*temp11;}

    if (temp1 > T1temp){
        u1temp = U1;
        T1temp = temp1;}}}
```

/ Routine that finds new $\mu_2(x_1, x_2, a)$ values from the 2nd expression of (3.4), given new $\mu_1(x_1, x_2, a)$ values */*

```
void optU2(int X1, int X2, int a, int U1)
{long double temp2;
long double cumprob2;
long double prob2;
long double temp22;

T2temp = – 10000;
for (U2 = – min(0, X2); U2 <= min(U2max, X2max – X2); U2++){
    Wmax1 = X1 + U1 – X1min;
    Wmax2 = X2 + U2 – X2min;
    Temp2 = (long double) –C2*U2;

    if (P[a] != 1){
        cumprob2 = 0.0;
        temp22 = 0.0;
        for (w = 0; w <= X2 + U2; w++ ){
            prob2 = p*(pow((1 – p), w));
            cumprob2 = cumprob2 + prob2;
            temp22 = temp22 + (prob2)*(r2*w – h2*(X2 + U2 – w) + V2[(X1 + U1) – X1min][(X2 + U2 –
                w) – X2min][max((a – 1), 0)]); }
        for (w = min(X2 + U2 + 1, Wmax2); w <= Wmax2; w++){
            prob2 = p*(pow((1 – p), w));
            cumprob2 = cumprob2 + prob2;
            temp22 = temp22 + (prob2)*(r2*w + V2[(X1 + U1) – X1min][(X2 + U2 – w) – X2min][min((a
                + 1), M)]);}

        temp22 = temp22 + (1 – cumprob2)*(r2*(X2 + U2 – X2min) + V2[(X1 + U1) – X1min][( X2min) –
            X2min][min((a + 1), M)]);
        temp2 = temp2 + (1 – P[a])*temp22; }

    if (P[a] != 0){
        cumprob2 = 0.0;
        temp22 = 0.0;
        for (w = 0; w <= (X1 + U1); w++ ){
            prob2 = p*(pow((1 – p), w));
            cumprob2 = cumprob2 + prob2;
            temp22 = temp22 + (prob2)*(V2[(X1 + U1 – w) – X1min][(X2 + U2) – X2min][min((a + 1),
                M)] – h2*max(0, ( X2 + U2))); }
```

```
        for (w = min(X1 + U1 + 1, Wmax1); w <= Wmax1; w++ ){
            prob2 = p*(pow((1 – p), w));
            cumprob2 = cumprob2 + prob2;
            temp22 = temp22 + (prob2)*(V2[(X1 + U1 – w) – X1min][(X2 + U2) – X2min][max((a – 1),
                0)] – h2*max(0, (X2 + U2)));}

        temp22 = temp22 + (1 – cumprob2)*(V2[(X1min) – X1min][(X2 + U2) – X2min][max((a – 1), 0)]
            – h2*max(0, ( X2 + U2)));
        temp2 = temp2 + (P[a])*temp22; }

    if (temp2 > T2temp){
    u2temp = U2;
    T2temp = temp2;}}}
```

/ Routine that finds new optimal control and differential profit mapping values from expression (3.5)*/*

```
void optU1U2(int x1, int x2, int a)
{int u2temp_old;
u2temp_old = X2max – X2 + 100;
u2temp = X2max – X2;

while (u2temp != u2temp_old){
    u2temp_old = u2temp;
    /* Given old μ2(x1, x2, a) values, find new μ1(x1, x2, a) values from the 1st expression of (3.4) */
    optU1(X1, X2, a, u2temp_old);
    /* Given new μ1(x1, x2, a) values, find new μ2(x1, x2, a) values from the 2nd expression of (3.4) */
    optU2(X1, X2, a, u1temp);}

UTable1[X1 – X1min][X2 – X2min][a] = u1temp;
UTable2[X1 – X1min][X2 – X2min][a] = u2temp;
TTable1[X1 – X1min][X2 – X2min][a] = T1temp;
TTable2[X1 – X1min][X2 – X2min][a] = T2temp;}
```

/ Routine that prints optimal order quantities $u_i^*(x_1, x_2, a) = \mu_i^*(x_1, x_2, a),\ i = 1, 2$ */*

```
void print_UTTables()
{int i, j, k;
fptr = fopen("results.txt", "w");

for (k = 0; k <= M; k++ ){
    fprintf(fptr, "Printing UTable1 for a = %d \n", k);
    for (i = X2max – X2min; i >= 0; i– – ){
        fprintf(fptr, "X2\t");
        for (j = 0; j <= X1max – X1min; j++ ){
            fprintf(fptr, " %d\t", UTable1[j][i][k]); }
        fprintf(fptr, "\n");}
```

```
    fprintf(fptr, "Printing UTable2 for a = %d \n", k);
        for (i = X2max – X2min; i >= 0; i– – ){
            fprintf(fptr, "X2\t");
            for (j = 0; j <= X1max – X1min; j++ ){
                fprintf(fptr, "%d\t", UTable2[j][i][k]);}
        fprintf(fptr, "\n");}}
fclose (fptr); }
```

Declaration Routine:

```
/* Routine with variable and parameter declarations */
#include <stdio.h>
#include <math.h>
#include <stdlib.h>
#include <time.h>

FILE *fptr;
/* State-space grid parameters */
int X1max = 10;
int X2max = 10;
int X1min = – 41;
int X2min = – 41;

#define N 53
#define Nss 52

#define p 0.35 /* Parameter ρ of the geometric distribution of customer demand */
#define M 3    /* Number of credibility levels */

/* reward and cost parameters */
#define h1 0.01
#define h2 0.01
#define r1 10
#define r2 10
#define C1 5
#define C2 5

/* Upper limit of control variables */
#define U1max 100
#define U2max 100

int w, U1, U2, u1temp, u2temp, X1, X2, a, Wmax1, Wmax2;

long double T1, T2, T1temp, T2temp, temp1, temp2;
long double V1diffmax, V2diffmax;
long double V1diffmin, V2diffmin;
long double Tempdiff1, Tempdiff2;
```

```
long double P[M + 1];
long double V1[Nss][Nss][M + 1] = {0};
long double V2[Nss][Nss][M + 1] = {0};
long double V1_old[Nss][Nss][M + 1];
long double V2_old[Nss][Nss][M + 1];
long double T_Table1[Nss][Nss][M + 1] = {0};
long double T_Table2[Nss][Nss][M + 1] = {0};
int U_Table1[Nss][Nss][M + 1] = {0};
int U_Table2[Nss][Nss][M + 1] = {0};
int X_1´ = 0;
int X_2´ = 0;
int a_1´ = 1;
int a_2´ = 0;
long double e = 0.00001;
```

Chapter 4 Do Stockouts Undermine Current Sales and Future Customer Demand?

The main goal of this chapter is to shed some light into the effect of stockouts on current sales and future customer demand in a wholesale business environment. To this end, we study the linkage between stockouts, customer service, current sales, and future demand, by performing a thorough statistical analysis of historical customer order and delivery data of a tool wholesaler and distributor over a period of four years. We hope that the results of this analysis will provide useful information to operations management (OM) researchers who wish to develop and analyze realistic models of supplier-customer behavior. Our analysis could also serve as an example for sales and inventory management practitioners who wish to perform a similar study on their own data. Another objective of this chapter is to statistically analyze the customer order data itself. Given the lack of reports on real customer demand data in the literature, this analysis may be of particular interest to inventory management researchers who wish to develop and analyze realistic models of customer demand.

Recently, there has been an increasing call for rigorous empirical research in OM. In contrast to other more mature management disciplines, OM has the least developed empirical knowledge base to draw upon in answering challenging questions. This may be due to at least two reasons. Firstly, empirical research involves a systematic derivation and analysis of data from direct or indirect observation, a job that most OM researchers are not well trained or interested in doing. Secondly, most companies that have the data are hesitant to share it with the rest of the world. We hope that this chapter will add an empirical contribution to the OM literature.

The remaining of this chapter is organized as follows. In Section 4.1, we discuss the procedure of collecting the order and delivery data and transforming it into meaningful variables that measure customer service, order fill rate, and the rate of future demand. In Section 4.2, we discuss the values of the sample means and coefficients of variation (CVs) of these variables. In Section 4.3, we perform a trend analysis of the customer order data. In Section 4.4, we identify the distributions of the order data, and in Section 4.5, we test for the existence of autocorrelations in the order data. In Section 4.6, we test for the existence of correlation between customer service and order fill rate, and in Section 4.7, we test for the existence of correlation between customer service and the rate of future demand. In Section 4.8, we try several commonly used nonlinear regression models between the variables measuring customer service and those measuring the rate of future demand. Finally, in Section 4.9, we summarize our findings and discuss their implications.

4.1 Data Collection

The company that provided the customer order and delivery data for this study was established as a retailer of ironware in Central Greece in 1922. Today, it is a large wholesaler and distributor of imported hand tools, hardware, industrial tools and equipment, electric power tools, accessories for power tools, welding machines and accessories, agricultural implements, and other similar products. The company imports and distributes products of many major European and Asian tool manufacturers in a very competitive environment. It is also an exclusive importer and distributor of a limited number of European and American tool manufacturers. The facilities of the company include a large central warehouse and two local retail shops that sell hand tools, industrial equipment, electric power tools and related equipment. The sales department of the company is staffed with twelve well-trained salespersons that travel in company owned cars to support over two thousand customers throughout the country, including numerous islands. The customers are retail shops and smaller distributors, i.e. the company operates in a B2B market.

Customers place their orders usually by toll-free phone or fax and sometimes by email, and expect their orders to be met immediately. Each order typically contains several items (SKUs) in different quantities and with different prices and is handled by the salesperson who has been assigned to the customer that placed the order. The items of the order that are in stock are delivered to the customer usually on the next working day. Same-day delivery is possible for orders that are placed before noon. The items that are out of stock are backordered. The backordered items must be delivered within a specific time frame set by the customer. If they are not delivered by this time frame, the order for these items is cancelled. If this time frame is zero, the order for items that are out of stock is immediately cancelled. Usually, an order is partially met in more than one delivery, and part of it may be cancelled.

The company keeps a record of every order, including the items that are cancelled. For each order, it also keeps a record of the delivery dates and the items delivered on those dates. From these records we extracted the order and delivery information for the nine most important customers of the company, for a period of four years that included 1043 working days, from January 1, 1999 to December 31, 2002. To simplify the analysis of the data, we aggregated all the items in each order and expressed each order and its deliveries in terms of their monetary values in €. More specifically, for each order *i* of each customer, we collected the following raw data:

a_i : arrival date of the order;

d_i : monetary value of the order that was placed, including the value of the items that were eventually cancelled from the order;

J_i : number of deliveries of the order;

$b_{j,i}$: delivery date of the *j*th delivery of the order, $j = 1, \ldots, J_i$;

$q_{j,i}$: monetary value of the *j*th delivery of the order, $j = 1, \ldots, J_i$;

$\kappa_{j,i} \equiv b_{j,i} - a_i$: delay in number of working days between the arrival date of the order and the *j*th delivery date of the order, $j = 1, \ldots, J_i$.

The main goal of our study was to examine if the customer service that the company provided to any particular order of a customer affected the fill rate of that order, i.e. the fraction of the order that was eventually delivered (not cancelled), as well as the rate of future orders of the same customer, and if so, how strong these effects were. To this end, we defined a set of variables to be used as measures of the customer service level, the order fill rate, and the rate of future orders. More specifically, for each order *i*, we defined the following variables as measures of the customer service level and the order fill rate, and computed their values:

$x_i \equiv 1 - \sum_{j=1}^{J_i} q_{j,i} \Big/ d_i$: fraction of the value of the order that was cancelled;

$k_i \equiv \kappa_{J_i,i}$: maximum delivery delay;

$f_i \equiv \sum_{j=1}^{J_i} \kappa_{j,i} \left(q_{j,i} / d_i \right) + 2 k_i x_i$: weighted sum of delivery delays.

In the above expression for f_i, the first term represents the weighted sum of the delivery delays, where the delay of each delivery is weighted by the fraction of the value of the order that was filled in this delivery. The second term represents a penalty for not delivering the cancelled part of the order weighted by the fraction of the value of the order that was cancelled. This penalty was chosen somewhat arbitrarily. We knew that it had to be larger than the maximum delivery delay, so we chose it to be twice the value of the maximum delivery delay. For each order i, we also defined the following variables as measures of the rate of future orders, and computed their values:

$e_i \equiv a_{i+1} - a_i$: number of working days until the arrival of the next order;

$h_i \equiv d_{i+1}$: monetary value of the next order placed.

We then entered all the above data into separate order and delivery data tables, one table for each customer. An example of such a table is Table 4-1, which shows the data for customer 4.

Table 4-1: Table of order and delivery data for customer 4

Order No.	Raw data												Service level and fill rate variables			Future order variables	
i	a_i	d_i	J_i	$b_{1,i}$	$q_{1,i}$	$b_{2,i}$	$q_{2,i}$	$b_{3,i}$	$q_{3,i}$	$\kappa_{1,i}$	$\kappa_{2,i}$	$\kappa_{3,i}$	x_i	k_i	f_i	e_i	h_i
1	1/29/99	614.91	2	2/3/99	449.17	2/6/99	165.74	-	-	3	5	-	0	5	3.54	9	574.15
2	2/11/99	574.15	2	2/12/99	369.8	2/26/99	204.35	-	-	1	11	-	0	11	4.56	5	1573.96
⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮
46	12/4/02	77.57	1	12/28/03	77.57	-	-	-	-	17			0	17	17	1	141.6
47	12/5/02	141.6	1	12/9/02	141.6	-	-	-	-	2			0	2	2	-	-

To get an idea of how the customer orders look like, Figure 4-1 shows a plot of the monetary value versus the arrival date of each order for customer 4.

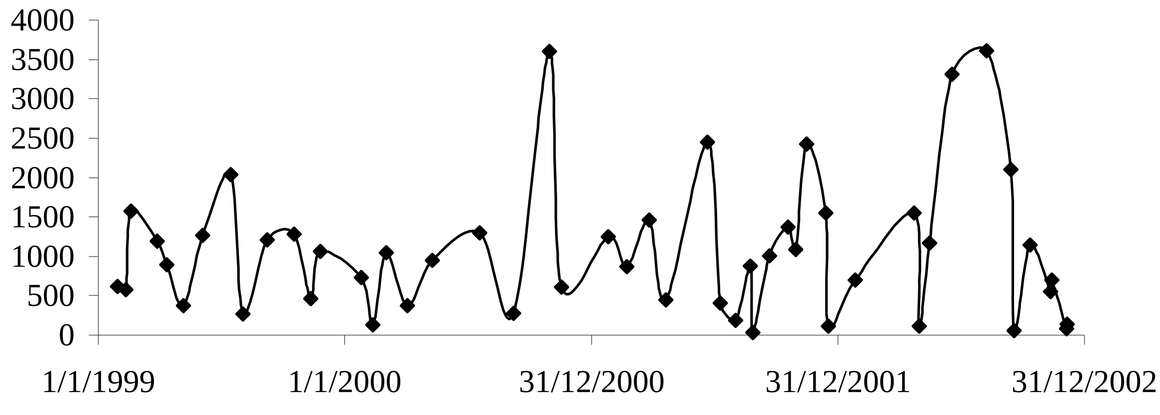

Figure 4-1: Monetary value in € versus arrival date for the orders of customer 4

Besides the tables that we created for each individual customer, we also created a table containing the order and delivery data for the ensemble of the customers. The way we created this table was by aggregating all the orders from different customers that arrived in any particular day into a single order whose monetary value was the sum of the monetary values of the individual customer orders in that day. The set of delivery dates of this aggregate order was the union of the delivery dates of the individual orders, the number of deliveries was the number of elements of the set of delivery dates, and the monetary value of each delivery date was the sum of the monetary values of the orders delivered to all the customers on that date.

An issue that emerged after having collected the data was how to treat outliers. An outlier is an observation that lies an abnormal distance from all the other observed values. In order to identify the outliers in our data we applied the box plot technique. The box plot is a graphical display that uses the median and the lower and upper quartiles, defined as the 25 and 75 percentile and denoted by $Q1$ and $Q2$, respectively, to describe the behavior of the data in the middle and the ends of the distributions. A box plot is constructed by drawing a box between the upper and lower quartiles with a solid line drawn across the box to locate the median. The following quantities (called fences) are needed for identifying extreme values in the tails of the distribution: the lower and upper inner fences, defined as $Q1 - 1.5IQ$ and $Q2 + 1.5IQ$ respectively, and the lower and upper outer fences, defined as $Q1 - 3IQ$ and $Q2 + 3IQ$, respectively, where IQ denotes the inter-quartile range and is defined as the difference $Q2 - Q1$. A point beyond an inner fence on either side is considered a mild outlier. A point beyond an outer fence is considered an extreme outlier.

Using the box plot technique, we scanned the order and delivery data table of each customer and we eliminated those rows (orders) i whose corresponding e_i or h_i values were found to be extreme outliers. This way, we eliminated approximately 10% of the rows of each table. Had we not eliminated these data, we would have ended up with: (1) biased or distorted estimates (e.g. sample means and CVs), (2) inflated sums of squares, which would make it difficult to partition sources of variation in the data into meaningful components, (3) and distorted p-values, which would make it hard to make inferences about statistical significance, or lack thereof. More importantly, had we not looked for indications that there was something unusual in the data we might have drawn false conclusions. More specifically, if we had included the outliers in our analysis, we risked having attributed extremely high values of e_i or h_i to the customer service level of order i, when such extreme values were most likely due to an extraordinary event that caused the customer to deviate from his normal behavior. For example, if a customer decided to buy an unusually expensive item in his next order, the monetary value of that order would be abnormally large; yet, this should not have been related to the service level he received. Similarly, if a customer suspended his business for Christmas or Easter holidays, or summer vacations, or some other reason, the number of days until the arrival of his next order, e_i, would be abnormally large; yet, this would have nothing to do with the service level that the customer received prior to the suspension of his business. Note that when we computed the values of variables $\kappa_{j,i}$ and e_i as the difference in number of working days between two dates, we excluded the weekend days between the two dates, but we did not exclude any other days during which the company was closed, because this information was not available.

4.2 Discussion of the Sample Means and CVs of the Data

From the data that we collected for each customer and for the ensemble of the customers, we computed the sample mean and CV of all the variables, after having eliminated the elements corresponding to extreme outliers. The results are shown in Table 4-2, where N denotes the sample size and μ_y and c_y denote the mean and CV of any given variable y, respectively.

Table 4-2: Sample mean and CV of the order and delivery variables

Statistic	Customer									Ensemble
	1	2	3	4	5	6	7	8	9	
N	80	53	121	42	145	59	247	48	41	631
μ_e	8.4000	17.9434	5.9256	18.5952	4.7415	15.7966	2.8300	18.9167	18.6829	1.5166
c_e	0.9012	0.7302	0.8228	0.6010	0.7475	0.6473	0.7489	0.7244	0.6554	0.5680
μ_h	509.867	1749.450	443.077	868.006	181.840	500.020	76.352	403.289	890.080	759.470
c_h	0.8279	0.9304	1.1447	0.7317	1.1420	0.6552	0.7363	0.9094	0.8939	1.2023
μ_x	0.0310	0.0543	0.0453	0.0803	0.0447	0.0473	0.0180	0.0271	0.0351	0.0419
c_x	3.1807	1.5816	2.8148	1.4645	2.2196	1.7937	3.7844	3.6385	2.5932	2.4038
μ_k	3.1875	13.3774	4.0331	14.1190	4.5714	5.0508	0.6761	9.2500	14.2927	6.4675
c_k	2.6324	1.1937	2.0283	1.0503	5.2273	1.9536	3.4418	1.4019	1.3330	2.4330
μ_f	2.9795	4.9908	2.9352	6.0301	2.5095	3.0252	1.7865	4.5409	4.6097	1.8724
c_f	2.2602	1.1002	2.0763	0.9478	1.9906	1.1370	5.9746	1.8950	1.9171	1.8213

From the sample means of the variables of each individual customer shown in Table 4-2, we can see that the customers exhibited different average ordering behaviors. As a result, they received different average levels of customer service, to which they responded appropriately. To help us distinguish these differences, we constructed Table 4-3, in which we arranged the customers in decreasing order of their sample mean values for several variables.

From Table 4-3, we can see that the customers who have larger μ_h values, i.e., who placed larger orders, have larger μ_k and μ_f values, i.e., faced larger maximum and average delivery delays. This is most likely due to the fact the larger the order value is, the larger the number of items in the order, and therefore the larger the probability that some of the items are out of stock. In response to the reduced customer service that they received, these customers generally had larger order cancellation percentages, μ_x, except for customer 9, who had a relatively low cancellation percentage. From Table 4-3, we can also see that the customers who placed larger orders, generally had larger μ_e values, i.e., ordered less frequently, except for customer 8, who placed smaller orders relatively infrequently.

Table 4-3: Ordering of customers based on the sample mean of several order and delivery variables

Statistic	Larger ⟵							⟶ Smaller
μ_e	9	4	8	2	6	1	3	5	7
μ_h	2	9	4	1	6	3	8	5	7
μ_x	4	2	6	3	5	9	1	8	7
μ_k	9	4	2	8	6	5	3	1	7
μ_f	4	2	9	8	6	1	3	5	7

From the sample CVs of the variables shown in Table 4-2, we can observe that different variables exhibited different levels of variability. Hopp and Spearman (2000) classify a random variable as having low variability, moderate variability, or high variability, if its CV is smaller than 0.75, between 0.75 and 1.33, or greater than 1.33, respectively. Using this classification, we can see from the data that the number of days until the arrival of the next order, e, had low to moderate variability for all the customers as well as the ensemble of the customers. Similarly, the monetary value of the next order, h, had moderate variability for all the customers as well as the ensemble of the customers, except for customer 6 who had low variability. All the other variables, which are related to the service that the company provided to the customers and to the customers' immediate response to that service, i.e., x, k, and f, had high variability for most of the customers as well as the ensemble of the customers, whereas for the rest of the customers they had moderate variability.

4.3 Trend Analysis of the Order Data

After having computed the sample means and CVs of the order and delivery data, we set out to explore if there was a trend in the order data. To this end, for each customer, we aggregated the monetary values of all the orders that arrived within each month and plotted them against time. The plots are shown in Figures 4-2 to 4-4. We then performed a linear regression of the monthly orders to find out if there is a trend in the demand during the entire 4-year period studied. The resulting trend in € per day and the R^2 coefficient of the regression analysis are shown in Table 4-4. In the same table, we also display the average monthly monetary value of all orders during any year Y, denoted by dm_Y, for each of the four years that we examined.

From the results in Table 4-4, it appears that there was a small positive trend in the demand for seven out of nine customers, and a very small negative trend for the remaining two customers. For most customers, the R^2 coefficient was extremely small, meaning that the linear trend model explains very little of the variability in the monthly orders. To a large extent, this is due to the fact that for most customers the increment in the average monthly demand from year to year changed dramatically in size and, even worse, in sign. The only customers that did not exhibit ups and downs in their average monthly demand from year to year were customers 1, 3, and 9. Not surprisingly, these customers had the highest R^2 coefficients.

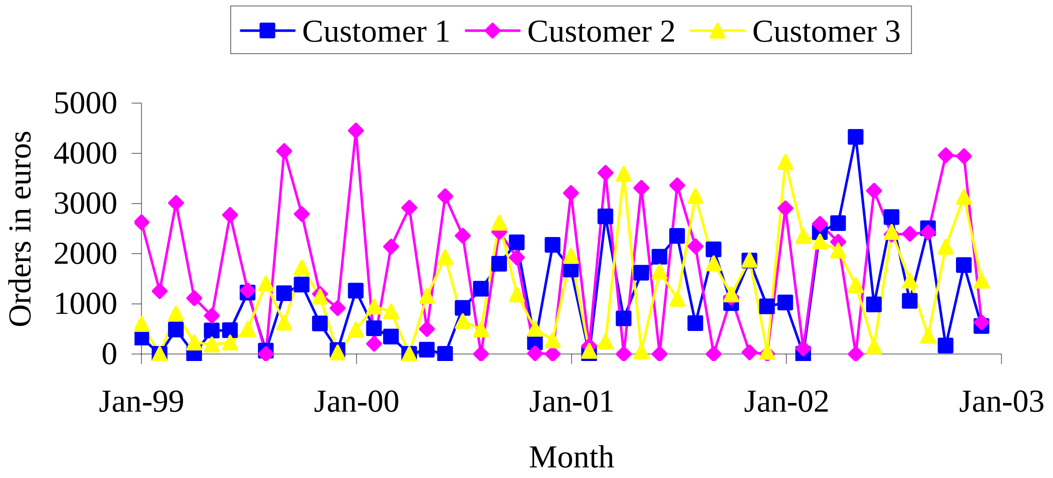

Figure 4-2: Monthly orders for customers 1-3

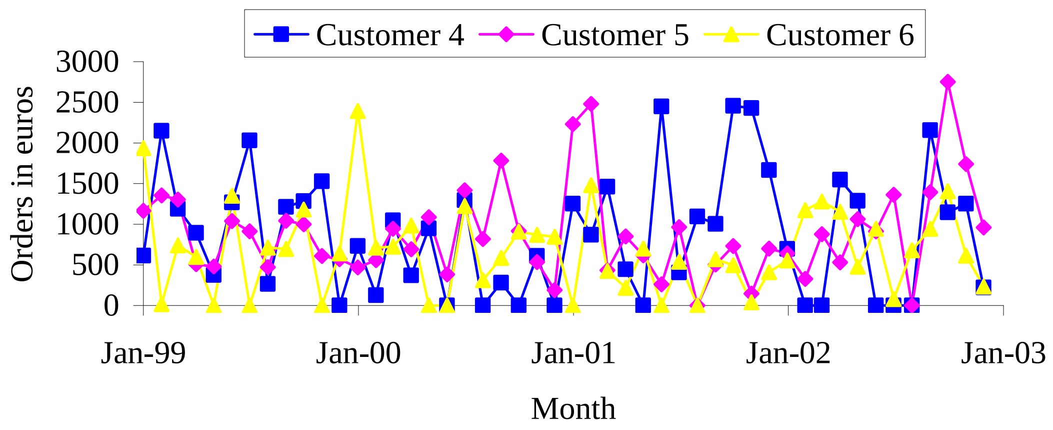

Figure 4-3: Monthly orders for customers 4-6

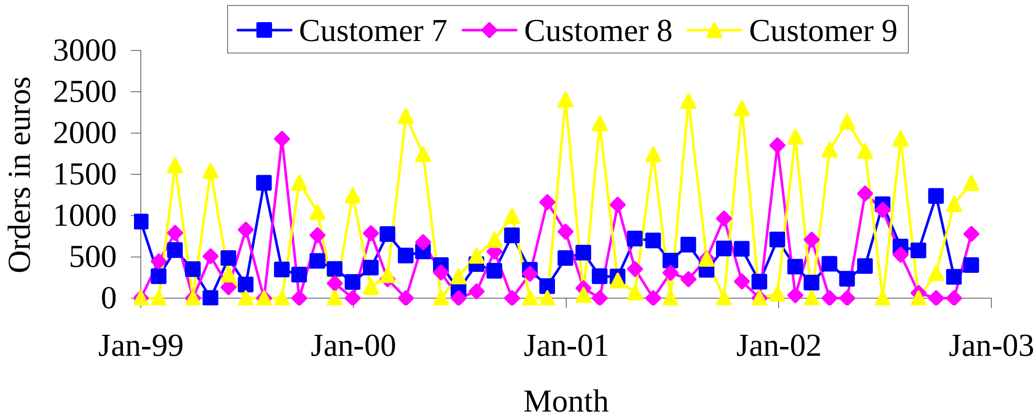

Figure 4-4: Monthly orders for customers 7-9

Table 4-4: Average monthly demand in € per month for each year, demand trend in € per day for the entire 4-year period, and corresponding R^2 coefficient

	Customer								
	1	2	3	4	5	6	7	8	9
dm_{1999}	525.6425	1812.551	625.0392	1067.327	872.045	656.305	465.9592	465.2525	490.2725
dm_{2000}	901.7273	1673.877	924.84	451.1033	816.3225	796.9983	408.3542	342.6708	674.0317
dm_{2001}	1587.33	1359.666	1471.438	1341.301	521.332	340.054	477.903	366.582	930.935
dm_{2002}	1678.218	2235.555	1918.204	692.3258	1046.498	795.2392	544.8942	525.6225	1042.158
4-year trend	1.0951	0.0893	1.1142	-0.0364	0.1048	-0.0658	0.0796	0.0487	0.4423
R^2	0.2259	0.0007	0.2191	0.0004	0.0058	0.0027	0.0143	0.0018	0.0470

4.4 Distribution Identification of the Order Variables

One of the main objectives of real data analysis is to determine the distributions of the variables that describe the physical data. Identifying candidate distributions is both an art and a science, as it requires an understanding of the underlying physical process, knowledge of the characteristics of the theoretical distributions, and a statistical analysis of the data. We studied the histograms and descriptive statistics of the historical data and fitted several candidate theoretical distributions for the order interarrival times and monetary values, *e* and *h*, respectively, for each individual customer and for the ensemble of the customers. For each candidate distribution, we used a least-squares fit followed by a probability plot (PP) in order to confirm the goodness of fit of that distribution. In a PP, the cumulative proportion for a variable is plotted against the cumulative proportion expected if the sample were from a specific theoretical distribution. If the sample is from the specific distribution, points will cluster around a straight line. In addition, we created detrended PPs that show the individual divergences between the observed and estimated cumulative values.

The results showed that the Weibull distribution provided the best or close to the best fit for every variable. This is not surprising given that the flexibility of the Weibull distribution allows it to fit many data sets. The parameters of the Weibull distribution are its *shape* and *scale*. The shape parameter of the Weibull distribution, denoted by β, provides insight into the behavior of the random variable of interest and in particular the shape of its hazard rate function in case the variable represents the time between two random events. The *hazard rate* function is a well-known function in reliability theory that provides an instantaneous (at time t) rate of occurrence of a random event such as a failure or, in our case, the arrival of a customer order. A value of $\beta > 1$ signifies an increasing hazard rate function, whereas a value of $\beta < 1$ signifies a decreasing hazard rate function. When $\beta = 1$, the hazard rate function is constant and the Weibull distribution is identical to the exponential distribution. When $1 < \beta < 2$, the hazard rate is increasing and concave. When $\beta = 2$, the hazard rate is increasing and linear. When $\beta > 2$, the hazard rate is increasing and convex. Finally, when $\beta < 1$, the Weibull distribution is similar in shape to the exponential, whereas when $\beta > 3$, the Weibull distribution is somewhat symmetrical like the normal distribution. When $1 < \beta < 3$, the Weibull distribution is skewed to the left. The scale parameter of the Weibull distribution, denoted by θ, influences both the mean and the spread of the distribution. As θ increases, the probability that the event will not occur at a given point in time increases, whereas the slope of the hazard rate decreases.

The parameters of the Weibull distribution for the variables e and h of each customer and the ensemble of the customers are shown in Table 4-5, where ι denotes the order of fit and is defined as the order of the maximum individual divergence between the observed and estimated cumulative values. An order of fit below 0.15 indicates a very good fit.

From the results in Table 4-5, we can see that the shape parameter β of the Weibull distribution of the number of days until the next order arrival, e, is between 1 and 2 for all the customers as well as for the ensemble of the customers. This means that the Weibull distribution is skewed to the left. It also means that the order interarrival times have an increasing and concave hazard rate, i.e. the longer the time since the last order arrival date, the larger the probability that the next order will arrive soon. This is natural, because as the time since the last order arrival date of any particular customer passes, this customer’s inventories are being depleted by his own customers and so the probability that he will soon place a replacement order increases. The fact that $\beta > 1$ for all the customers and for the ensemble of the customers also means that the interarrival time distributions deviate from the exponential distribution, for which $\beta = 1$, although not dramatically, since for five out of nine customers, β is below 1.2, and for the remaining four customers it is between 1.2 and 1.4. The results in Table 4-5 also show that the shape parameter β of the Weibull distribution of the monetary values of the orders, *h*, is between 1 and 2 for seven out of nine customers. This means that the Weibull distribution is skewed to the left for these customers. For the remaining two customers as well as for the ensemble of the customers β is less that 1, which means that the Weibull distribution is similar in shape to the exponential.

Table 4-5: Weibull parameters of the interarrival times and monetary values of the orders

Var	Par	Customer									Ensemble
		1	2	3	4	5	6	7	8	9	
	θ	8.6553	19.76	6.2803	21.6131	5.174	18.078	3.08	20.7746	20.9895	1.6984
e	β	1.0925	1.1	1.1713	1.3460	1.364	1.186	1.312	1.1777	1.2404	1.6969
	ι	0.05	0.08	0.05	0.10	0.03	0.08	0.05	0.12	0.11	0.04
	θ	540.394	1624.712	207.902	957.475	191.067	566.257	83.272	415.053	885.922	663.013
h	β	1.218	0.6932	1.372	1.1386	1.509	1.2533	1.631	1.1636	0.8895	0.8979
	ι	0.05	0.10	0.07	0.11	0.06	0.11	0.07	0.07	0.125	0.06

Finally, from Table 4-5, we can see that the scale parameter θ of the Weibull distribution of both variables *e* and *h* is close to the average values of these variables, μ_e and μ_h, which are displayed in Table 4-2.

4.5 Determination of Autocorrelation in the Order Data

Many analytical models of inventory management systems assume that the customer order interarrival times as well as the order sizes are independent random variables. In this section we test the validity of this assumption for the historical customer order data that we collected by testing for the existence of autocorrelation in that data. Lack of autocorrelation is necessary but not sufficient to show that successive observations of a random variable are independent. For the purposes of practical inventory management, however, testing for autocorrelation should suffice as an indication of independence.

Using (auto) regression analysis, we calculated the autocorrelation coefficients for lags ranging from 1 to 10, and we performed the Durbin-Watson test for addressing the significance of the lag – 1 autocorrelation, for the times between consecutive customer orders, *e*, and the monetary values of each order, *h*, for each individual customer and the ensemble of the customers. The results are shown in Tables 4-6 and 4-7, for *e* and *h*, respectively. Each table lists the lag – 1 autocorrelation coefficient, the maximum autocorrelation coefficient for any lag ranging from 1 to 10, and the lag corresponding to that maximum, which are labeled: lag – 1 *r*, max |*r*|, and lag for max |*r*|, respectively. Each table also lists the Durbin-Watson statistic *D* and 4 – *D*. One of the assumptions of regression analysis is that the residuals for consecutive observations are uncorrelated. If this is true, the expected value of the Durbin-Watson statistic *D* is 2. Values less than 2 indicate positive autocorrelation, and values greater than 2 indicate negative autocorrelation.

Table 4-6: Autocorrelation and Durbin-Watson test for the customer order interarrival times *e*

	Customer									
	1	2	3	4	5	6	7	8	9	Ensemble
	Autocorrelation									
Lag – 1 *r*	-0.027	-0.037	0.209	-0.084	-0.117	0.175	0.178	-0.011	-0.208	0.18
Max |*r*|	0.333	0.143	0.252	0.195	0.163	0.183	0.178	0.252	0.195	0.22
Lag for Max |*r*|	12	9	12	2	2	3	1	1	7	9
	Durbin-Watson Statistic									
D	1.986	2.035	2.03	1.952	1.958	1.963	1.99	1.976	1.951	2.049
4 - *D*	2.014	1.965	1.97	2.048	2.042	2.037	2.01	2.024	2.049	1.951
	Durbin-Watson 0.01 Test Bounds									
$D_{0.01,L}$	1.47	1.32	1.52	1.25	1.52	1.38	1.52	1.32	1.25	1.52
$D_{0.01,U}$	1.52	1.4	1.56	1.34	1.56	1.45	1.56	1.4	1.34	1.56
	Conclusion for H_0: $\rho = 0$, and									
H_1: $\rho > 0$	$\rho = 0$	$\rho = 0$	$\rho = 0$	$\rho = 0$	$\rho = 0$	$\rho = 0$	$\rho = 0$	$\rho = 0$	$\rho = 0$	$\rho = 0$
H_1: $\rho < 0$	$\rho = 0$	$\rho = 0$	$\rho = 0$	$\rho = 0$	$\rho = 0$	$\rho = 0$	$\rho = 0$	$\rho = 0$	$\rho = 0$	$\rho = 0$

From Tables 4-6 and 4-7, we can see that the Durbin-Watson statistic *D* is very close to 2 for all the *e* and *h* data. This means that the lag – 1 autocorrelation in the data is very small. The significance of the lag – 1 autocorrelation can be addressed with the Durbin-Watson test. This test compares *D* with upper and lower bounds $D_{\alpha,U}$ and $D_{\alpha,L}$ for a given significance level α. For the positive autocorrelation hypothesis ($\rho > 0$), if $D < D_{\alpha,L}$, we conclude that there is positive autocorrelation and if $D > D_{\alpha,U}$, we conclude that there is not. If $D_{\alpha,L} < D < D_{\alpha,U}$, the test is inconclusive. For the negative autocorrelation hypothesis ($\rho < 0$), if $4 – D < D_{\alpha,L}$, we conclude that there is negative autocorrelation, and if $4 - D > D_{\alpha,U}$, we conclude that there is not. If $D_{\alpha,L} < 4 – D < D_{\alpha,U}$, the test is inconclusive (e.g., see Hines and Montgomery, 1990). From Tables 4-6 and 4-7, it can be seen that the conclusion of all the tests is that there is neither positive nor negative lag – 1 autocorrelation at the 0.01 significance level in any of the data. Therefore, assuming independence appears valid for all practical purposes for both *e* and *h*.

Table 4-7: Autocorrelation and Durbin-Watson test for the customer order monetary values *h*

	Customer									Ensemble
	1	2	3	4	5	6	7	8	9	
	Autocorrelation									
Lag – 1 r	-0.11	-0.259	-0.012	0.097	0.023	0.119	-0.025	-0.123	-0.055	0.033
Max $\|r\|$	0.163	0.222	0.276	0.19	0.155	0.238	0.195	0.276	0.248	0.118
Lag for Max $\|r\|$	8	3	4	8	7	2	12	3	2	16
	Durbin-Watson Statistic									
D	2.014	1.987	1.923	1.99	1.995	1.908	1.997	2.075	1.966	2
4 - D	1.986	2.013	2.077	2.01	2.005	2.092	2.003	1.925	2.034	2
	Durbin-Watson 0.01 Test Bounds									
$D_{0.01,L}$	1.47	1.32	1.52	1.25	1.52	1.38	1.52	1.32	1.25	1.52
$D_{0.01,U}$	1.52	1.4	1.56	1.34	1.56	1.45	1.56	1.4	1.34	1.56
	Conclusion for H_0: $\rho = 0$, and									
H_1: $\rho > 0$	$\rho = 0$	$\rho = 0$	$\rho = 0$	$\rho = 0$	$\rho = 0$	$\rho = 0$	$\rho = 0$	$\rho = 0$	$\rho = 0$	$\rho = 0$
H_1: $\rho < 0$	$\rho = 0$	$\rho = 0$	$\rho = 0$	$\rho = 0$	$\rho = 0$	$\rho = 0$	$\rho = 0$	$\rho = 0$	$\rho = 0$	$\rho = 0$

4.6 Determination of Correlation between Customer Service and Order Fill Rate

In Section 4.2, we conjectured that customers who face larger company delivery delays respond to these delays with larger order cancellation percentages. This conjecture was based on a rough comparison between the mean value of the maximum delivery delay *k* and the order cancellation percentage *x* among different customers. To better support and refine this conjecture, we tested for the existence of correlation between customer service and the order fill rate for each customer separately as well as for the ensemble of the customers. More specifically, we examined if the maximum delivery delay, *k*, which measures customer service, was statistically correlated with the fraction of the value of the order that was cancelled, *x*, which determines the order fill rate.

To test the existence of correlation between two random variables one needs to compute the correlation coefficient of the two variables. Correlation coefficients range in value from –1, indicating a perfect negative relationship, to +1, indicating a perfect positive relationship. There are several correlation coefficient definitions. The most common is Pearson's correlation coefficient which measures the linear association between two variables and is used if the variables are normally distributed. If two variables are not normally distributed or if their relationship is not linear, Pearson's correlation coefficient is not an appropriate statistic for measuring their association.

Scatter plots of the *k* and *x* variables revealed that the distributions of both these variables were significantly skewed to the left and hence were far from being normal. For this reason, Pearson’s correlation coefficient would be an inappropriate statistic to use. A more appropriate statistic is Spearman's ρ correlation coefficient, which measures the rank-order association between two variables and works regardless of the distributions of the variables. With this in mind, we computed Spearman's correlation coefficient ρ with its two-tailed significance level p for variable k and x, for each customer and for the ensemble of the customers. The results are shown in Table 4-8, where the correlations that are significant at a 0.05 level are marked with one asterisk, while those that are significant at a 0.01 level are marked with two asterisks.

Table 4-8: Spearman’s ρ correlation coefficient and corresponding two tailed significance level p regarding the correlation between variables k and x

	Customer									Ensemble
	1	2	3	4	5	6	7	8	9	
ρ	0.115	**0.221***	**0.491****	0.087	**0.413****	**0.257***	**0.153***	-0.077	0.09	**0.348****
p	0.308	**0.05**	**0.000**	0.585	**0.000**	**0.05**	**0.016**	0.603	0.575	**0.000**

From the results displayed in Table 4-8, we can see that for five customers, namely customers 2, 3, 5, 6, and 7, as well as for the ensemble of the customers, there is a statistically significant positive correlation between k and x either at the 0.01 or the 0.05 level, since the corresponding p-value is smaller than 0.01 or 0.05, respectively. The existence of these correlations indicates that when customers 2, 3, 5, 6, and 7, face larger company delivery delays, they respond with larger order cancellation percentages.

For the remaining four customers, namely customers 1, 4, 8, and 9, Spearman’s ρ coefficient is positive (except in the case of customer 8 where it is slightly negative) but not significantly different from zero at the 0.01 or even the 0.05 level, because the corresponding p-value is greater than 0.01 or 0.05, respectively. For these customers, therefore, there is no statistically significant evidence that the company’s delivery delays affect their order fill rate.

4.7 Determination of Correlation between Customer Service and Future Customer Orders

One of the main goals of our study was to examine if the poor service that a customer receives due to a stockout affects the rate of his future orders. To this end, we tested for the existence of correlation between customer service and future customer orders from the data that we collected. More specifically, we examined if the variables that measure the severity of stockouts, which we call *independent* variables, were statistically correlated with the variables that measure the change in the rate of future customer orders, which we call *dependent* variables.

The independent variables that measure the severity of a stockout faced by any particular order i of any particular customer are x_i, k_i, and f_i. The dependent variables that measure the change in the rate of future customer orders following order i are e_i and h_i. Intuition suggests that a drop in the rate of future customer orders may be affected not only by the most recent stockout experienced by a customer but by previous stockouts as well, although the effect of older stockouts on the drop in future customer demands should be less intense than the effect of more recent stockouts. In order to test the hypothesis that the drop – if any – in the rate of future customer orders due to the loss of customer goodwill is a phenomenon that is cumulative over time but at the same time customers are forgetting and forgiving as time passes, we introduced three new sets of variables, which were defined as the exponentially smoothed versions of the three original independent variables, x_i, k_i, and f_i. In each new variable, the severity of the stockout faced by order i is measured by weighing the current value as well as all the previous values of the respective variable with geometrically decreasing weights as we go back in time. More specifically, the exponentially smoothed versions of the independent variables were defined as follows:

$$X_i^\alpha \equiv \alpha x_i + (1-\alpha) X_{i-1}^\alpha$$

$$K_i^\alpha \equiv \alpha k_i + (1-\alpha) K_{i-1}^\alpha$$

$$F_i^\alpha \equiv \alpha f_i + (1-\alpha) F_{i-1}^\alpha$$

where α is the smoothing factor. Note that as α tends to 1, more and more weight is being placed on the more recent value of the independent variable, whereas as α tends to 0, more and more weight is being placed on past values of the independent variable. In this study, we considered four values for α, namely, 0.2, 0.4, 0.6, 0.8, and 1.

In Section 4.4, we saw that the distributions of the dependent variables e and h are skewed to the left and hence are far from being normal. For this reason, we argued that Pearson's correlation coefficient would be an inappropriate statistic to use in order to test the correlation between the independent and the dependent variables. A more appropriate statistic is Spearman's ρ correlation coefficient. With this in mind, we computed Spearman's correlation coefficient ρ with its two-tailed significance level p for each pair of independent variables X^{α}, K^{α}, and F^{α}, for α = 0.2, 0.4, 0.6, 0.8, and 1, and dependent variables e and h. The results are shown in Tables 4-9 and 4-10, where the correlations that are significant at a 0.05 level are marked with one asterisk, while those that are significant at a 0.01 level are marked with two asterisks.

From the results displayed in Tables 4-9 and 4-10, we can see that for seven out of nine customers at least one of the independent variables is statistically correlated with at least one of the dependent variables. This supports the allegation that stockouts adversely affect the rate of future customer orders, at least for most customers. Moreover, for the majority of the cases where there is a statistically significant correlation between an independent and a dependent variable, the corresponding Spearman’s ρ correlation coefficient is below 0.4, indicating that this correlation is not very strong.

From the results shown in Tables 4–9 and 4–10, we can see that for all the cases where the correlation between an independent variable and variable e is significant, this correlation is positive (see Table 4-9). This is in line with intuition, which suggests that the larger the value of the independent variable is, the lower the service level, and hence the longer the time until the next order, e. On the other hand, for all the cases where the correlation between an independent variable and variable h is significant, this correlation is negative (see Table 4-10). This is also intuitively reasonable, because the larger the value of the independent variable is, the lower the service level, and hence the smaller the value of the next order, h.

From Tables 4–9 and 4–10, we can also observe that the pairs of independent and dependent variables that have statistically significant correlations vary from customer to customer. This suggests that different customers respond differently to stockouts as far as their future orders are concerned. To see this, let us look in detail at the pairs of variables that have statistically significant correlation coefficients, for each customer separately.

Customers 1 and *5* are the only customers that exhibit no statistically significant correlation between any independent and any dependent variable. This means that there is no statistical evidence that customers 1 and 5 change the rate of their future orders in response to stockouts. This may be due to a number of factors that we did not take into account in this study, such as the size and level of sophistication of each customer, special pricing contracts between the company and its customers, the geographical proximity of each customer to the company, etc.

Table 4-9: Spearman's ρ correlation coefficient and corresponding two tailed significance level p regarding the correlation between each independent variable and variable e

		Customer								
Ind. var.		1	2	3	4	5	6	7	8	9
X^{1}	ρ	0.005	0.045	.266**	0.008	-0.021	-0.121	.165**	0.277	0.086
	p	0.962	0.748	0.003	0.958	0.803	0.363	0.01	0.057	0.593
$X^{0.8}$	ρ	0.041	0.023	.249**	-0.049	0.008	-0.085	0.053	0.146	0.191
	p	0.715	0.868	0.006	0.756	0.924	0.521	0.405	0.321	0.232
$X^{0.6}$	ρ	0.042	0.109	.230*	-0.157	0.031	-0.055	0.026	0.139	0.203
	p	0.71	0.439	0.011	0.32	0.707	0.677	0.688	0.347	0.204
$X^{0.4}$	ρ	0.062	0.171	.200*	0.24	0.05	-0.018	-0.02	0.118	0.185
	p	0.586	0.221	0.028	0.126	0.55	0.889	0.759	0.424	0.246
$X^{0.2}$	ρ	0.089	0.197	.184*	.383*	0.098	0.004	-0.102	0.056	0.158
	p	0.435	0.157	0.044	0.012	0.235	0.975	0.111	0.706	0.323
K^{1}	ρ	0.19	0.12	0.159	0.194	0.089	.306*	.223**	0.284	0.237
	p	0.091	0.392	0.081	0.219	0.285	0.018	0	0.05	0.136
$K^{0.8}$	ρ	0.193	0.22	0.1	0.262	0.084	.304*	.179**	0.169	0.264
	p	0.086	0.114	0.275	0.093	0.312	0.019	0.005	0.251	0.096
$K^{0.6}$	ρ	0.201	0.248	0.101	0.278	0.09	.295*	.167**	0.158	0.234
	p	0.074	0.073	0.268	0.074	0.28	0.023	0.008	0.283	0.142
$K^{0.4}$	ρ	0.16	.291*	0.123	0.294	0.11	.281*	.135*	0.129	0.17
	p	0.157	0.034	0.178	0.059	0.183	0.031	0.034	0.382	0.287
$K^{0.2}$	ρ	0.151	0.237	.204*	0.276	0.132	0.206	.128*	0.086	0.058
	p	0.18	0.088	0.024	0.077	0.111	0.118	0.044	0.562	0.72
F^{1}	ρ	0.113	0.13	.261**	0.096	-0.005	0.176	.268**	.298*	0.132
	p	0.32	0.352	0.004	0.546	0.954	0.181	0	0.04	0.409
$F^{0.8}$	ρ	0.126	0.114	.256**	0.072	0.03	0.192	.194**	0.262	0.215
	p	0.264	0.415	0.005	0.649	0.72	0.145	0.002	0.072	0.176
$F^{0.6}$	ρ	0.142	0.111	.239**	0.017	0.033	0.2	.141*	0.23	0.276
	p	0.208	0.427	0.008	0.914	0.689	0.128	0.027	0.117	0.081
$F^{0.4}$	ρ	0.142	0.111	.244**	-0.041	0.056	0.144	0.068	0.168	0.308
	p	0.208	0.43	0.007	0.796	0.503	0.276	0.288	0.253	0.05
$F^{0.2}$	ρ	0.164	0.029	.214*	-0.031	0.094	0.072	-0.01	0.118	0.296
	p	0.147	0.835	0.018	0.843	0.255	0.588	0.872	0.424	0.06

Table 4-10: Spearman's ρ correlation coefficient and corresponding two tailed significance level p regarding the correlation between each independent variable and variable h

		Customer								
Ind. var.		1	2	3	4	5	6	7	8	9
X^1	ρ	0.158	-0.147	-0.173	-0.27	-0.041	-0.199	0.009	**-.354***	-0.137
	p	0.161	0.294	0.058	0.083	0.618	0.13	0.893	0.014	0.394
$X^{0.8}$	ρ	0.175	-0.139	-0.116	**-.323***	-0.035	-0.169	-0.025	-0.023	**-.379***
	p	0.121	0.322	0.205	0.037	0.671	0.2	0.697	0.876	0.015
$X^{0.6}$	ρ	0.179	-0.103	-0.102	**-.369***	-0.019	-0.15	-0.019	-0.035	**-.389***
	p	0.111	0.463	0.267	0.016	0.818	0.256	0.764	0.811	0.012
$X^{0.4}$	ρ	0.163	-0.052	-0.08	**-.305***	0.005	-0.089	-0.009	-0.021	**-.340***
	p	0.149	0.712	0.383	0.05	0.948	0.502	0.891	0.889	0.029
$X^{0.2}$	ρ	0.162	-0.088	-0.053	-0.291	0.035	-0.004	0.006	-0.039	-0.277
	p	0.15	0.532	0.563	0.062	0.675	0.977	0.927	0.793	0.08
K^1	ρ	0.026	-0.067	**-.212***	-0.151	-0.134	-0.085	0.01	-0.104	-0.101
	p	0.817	0.636	0.02	0.341	0.104	0.523	0.878	0.48	0.529
$K^{0.8}$	ρ	-0.004	0.003	**-.181***	-0.104	-0.119	-0.159	-0.063	-0.082	-0.108
	p	0.972	0.984	0.047	0.514	0.152	0.229	0.327	0.578	0.502
$K^{0.6}$	ρ	-0.023	0.033	-0.166	-0.106	-0.121	-0.196	-0.068	-0.059	-0.116
	p	0.836	0.814	0.069	0.504	0.145	0.136	0.286	0.692	0.469
$K^{0.4}$	ρ	-0.066	0.023	-0.161	-0.098	-0.117	-0.218	-0.088	-0.032	-0.183
	p	0.558	0.868	0.078	0.538	0.157	0.097	0.167	0.828	0.252
$K^{0.2}$	ρ	-0.047	0.009	-0.157	-0.152	-0.094	-0.145	-0.075	-0.021	-0.189
	p	0.682	0.949	0.086	0.335	0.26	0.273	0.239	0.888	0.237
F^1	ρ	0.134	0.035	**-.215***	-0.064	-0.094	-0.095	0.051	-0.125	-0.026
	p	0.235	0.806	0.018	0.687	0.259	0.475	0.424	0.396	0.871
$F^{0.8}$	ρ	0.156	0.022	**-.199***	-0.091	-0.078	-0.075	0.057	-0.12	0.113
	p	0.167	0.874	0.029	0.565	0.347	0.573	0.374	0.416	0.482
$F^{0.6}$	ρ	0.122	-0.011	**-.183***	-0.131	-0.074	-0.067	0.059	-0.049	0.129
	p	0.281	0.938	0.045	0.407	0.375	0.616	0.357	0.743	0.42
$F^{0.4}$	ρ	0.09	-0.025	-0.144	-0.107	-0.051	-0.057	0.068	-0.042	0.156
	p	0.425	0.858	0.114	0.502	0.537	0.669	0.29	0.775	0.329
$F^{0.2}$	ρ	0.092	-0.073	-0.102	-0.153	-0.027	-0.108	0.102	-0.024	0.199
	p	0.415	0.603	0.264	0.334	0.746	0.414	0.109	0.869	0.213

For *customers 2* and *6*, the only statistically significant correlation that exists is between variables K^{α} and e. This means that there is statistical evidence that when customer 2 or 6 faces a long maximum delivery delay following a stockout, he extends the number of days until his next order. Moreover, customer 2 appears to have a fairly long memory of the disservice associated with the long maximum delivery delay, since the smoothing factor for which the correlation is statistically significant is much smaller that one (α = 0.4). For customer 6, on the other hand, the smoothing factor for which the correlation coefficient is the highest is one. This means that for customer 6, the correlation between the two variables is the highest if we assume that his behavior regarding future orders is only affected by the most recent stockout and not by past stockouts. The fact that there is no statistically significant correlation between the variables of the pairs (X^{α}, e) and (F^{α}, e) means that the rate of future orders of customers 2 and 6 is not affected by the fraction of the order that is cancelled or the weighted average delivery delay following a stockout. Similarly, the fact that there is no statistically significant correlation between any independent variable and h, means that there is no statistical evidence that customers 2 and 6 change the monetary value of their next order following a stockout.

For *customer 3*, there is a statistically significant correlation between every independent and every dependent variable, except for the pair of variables (X^{α}, h). This means that there is statistical evidence that customer 3 extends the number of days until his next order *and* lowers the monetary value of his next order following a stockout. The correlations which seem to be stronger are those between the variables of the pairs (X^{α}, e) and (F^{α}, e). This means that for customer 3, it is the fraction of the order that is cancelled following a stockout and the weighted average delivery delay rather than the maximum delivery delay that mostly affect the rate of his future orders. A closer look at the results reveals that for all the pairs of variables that exhibit statistically significant correlations, the smoothing factor for which the correlation coefficient is the highest is one, except for the pair (K^{α}, e). This means that for customer 3, the correlation between the two variables is the highest if we assume that his behavior regarding future orders is affected only by the most recent stockout and not by previous stockouts.

For *customer 4*, there are statistically significant correlations between the variables of the pairs (X^{α}, e) and (X^{α}, h). Therefore, there is statistical evidence that when customer 4 cancels a large fraction of his order following a stockout, he extends the number of days until his next order *and* he lowers the monetary value of his next order. Moreover, he appears to have a fairly long memory of the disservice associated with the stockout, since the smoothing factor for which the statistically significant correlation coefficient is higher is much smaller than one ($\alpha = 0.2$ for the pair (X^{α}, e) and $\alpha = 0.6$ for the pair (X^{α}, h)).

For *customer 7*, there is a statistically significant correlation between every independent variable and e, but no correlation between any independent variable and h. This means that there is statistical evidence that customer 7 extends the number of days until his next order but does not lower the monetary value of his next order following a stockout. A close examination of the results reveals that wherever there is a statistically significant correlation, the higher the smoothing factor is, the higher the correlation coefficient. This implies that for customer 7, the correlations between the independent variables and e are the highest if we assume that his behavior regarding future orders is affected only by the most recent stockout and not by previous stockouts.

For *customer 8*, there are statistically significant correlations between the variables of the pairs (F^{α}, e) and (X^{α}, h). This means that there is statistical evidence that when customer 8 faces a long weighted average delivery delay following a stockout, he extends the number of days until his next order. In addition, when he cancels a large fraction of his order following a stockout, he lowers the monetary value of his next order. A close examination of the results reveals that the smoothing factor for which the correlations are statistically significant is one. This means that the behavior of customer 8 regarding future orders is affected only by the most recent stockout and not by previous stockouts.

Finally, *customer 9* is the only customer that exhibits a statistically significant correlation between an independent variable, namely X^{α}, and h, but no correlation between any independent variable and e. This means that there is statistical evidence that when customer 9 cancels a large fraction of his order following a stockout, he reduces the monetary value of his next order but he does not extend the number of days until his next order. Moreover, he appears to have a fairly long memory of the disservice associated with the stockout, since the smoothing factor for which the statistically significant correlation coefficient is higher is smaller than one ($\alpha = 0.6$).

To summarize, for two out of nine customers there is no statistical evidence that stockouts affect the rate of their future orders. For the remaining seven customers, there is statistically significant evidence that stockouts affect either the time until their next order, e, or the monetary value of their next order, h, or both. More specifically, for three customers, stockouts affect both e and h, for three other customers they affect only e, and for one customer they affect only h; therefore, it seems that most customers who experience stockout reduce the frequency of their future orders and some lower the amount that they order.

Table 4-11 summarizes the results presented above by showing for each pair of independent and dependent variables and for each customer, the smoothing factor of the independent variable for which the statistically significant correlation (if any) between the two variables is the highest.

From Table 4-11, we can see that in the ten cases in which there is a statistically significant correlation between an independent variable and variable e, this independent variable is X^{α}, K^{α}, and F^{a} in 3, 4, and 3 cases, respectively. This means that the number of days until the next order is almost equally affected by the fraction of the current and past orders that were cancelled and the maximum as well as the weighed average delivery delay of the current and past orders. Moreover, in the five cases where there is statistically significant correlation between an independent variable and variable h, this independent variable is X^{α}, K^{α}, and F^{a} in 3, 1, and 1 case, respectively. This means that in most cases the monetary value of the next order is mostly affected by the fraction of the current and past orders that were cancelled.

Table 4-11: Smoothing factor for which the statistically significant correlation between the correlated variables is the highest

	Customer								
Correlated vars.	1	2	3	4	5	6	7	8	9
(e, X^{α})			1	0.2			1		
(e, K^{α})		0.4	0.2			1	1		
(e, F^{α})			1				1	1	
(h, X^{α})				0.6				1	0.6
(h, K^{α})			1						
(h, F^{α})			1						

4.8 Regression Analysis

After having verified the existence of statistically significant correlations between several independent and dependent variables, we used nonlinear regression to find what – if any – is the relationship between each set of dependent and independent variables that exhibited a statistically significant correlation.

The first step in finding accurate and reliable nonlinear regression models between the independent and dependent variables that exhibited significant correlations was to create scatter plots, which could help us recognize the nature of their relationship. Unfortunately, the scatter plots did not reveal any evident relationship in any of the pairs of variables examined. The next step was to produce curve estimation regression statistics and related plots for several different commonly used curve estimation regression models shown in Table 4-12, where x is the independent variable and $\hat{y}$ is the estimate of the dependent variable. Table 4-13 shows the best fitted regression model and its corresponding characteristics, namely, the R^2 coefficient, the significance level, and the values of the equation coefficients, for the pairs of variables with the highest statistically significant correlation coefficients shown in Table 4-11.

Table 4-12: Common curve estimation regression models

Regression model	Curve expression
Logarithmic	$\hat{y} = b_0 + b_1 \ln(x)$
Inverse	$\hat{y} = b_0 + b_1 / x$
Cubic	$\hat{y} = b_0 + b_1 x + b_2 x^2 + b_3 x^3$
Power	$\hat{y} = b_0 x^{b_1}$
Compound	$\hat{y} = b_0 b_1^x$
S-curve	$\hat{y} = e^{(b_0 + b_1/x)}$
Exponential	$\hat{y} = b_0 e^{b_1 x}$

Table 4-13: Best fitted regression model for the pairs of variables with the highest statistically significant correlation coefficients

Customer	Dep. var.	Ind. var.	Regression Model	R^2	Significance Level	b_0	b_1	b_2	b_3
2	*e*	$K^{0.4}$	Cubic	0.145	0.05	7.64	1.55	0.089	0.0022
3	*e*	X^1	Cubic	0.071	0.035	5.358	0.479	-0.016	0.000
	e	$K^{0.2}$	Cubic	0.111	0.003	3.94	0.44	-0.02	0.0026
	e	F^1	Cubic	0.078	0.023	4.879	0.799	-0.042	0.001
	h	K^1	Exponential	0.045	0.02	233.94	-0.036		
	h	F^1	Exponential	0.064	0.005	245.877	-0.057		
4	*e*	$X^{0.2}$	Exponential	0.220	0.002	31.82	-0.11		
	h	$X^{0.6}$	Exponential	0.246	0.001	1058.71	-0.07		
6	*e*	K^1	Cubic	0.154	0.026	11.891	2.342	-0.124	0.002
7	*e*	X^1	*S*-curve	0.035	0.003	1.160	$-4.3\cdot10^{-6}$		
	e	K^1	*S*-curve	0.05	0.000	1.061	$-3.7\cdot10^{-6}$		
	e	F^1	*S*-curve	0.07	0.000	1.068	$-4.2\cdot10^{-6}$		
8	*e*	F^1	Cubic	0.138	0.08	12.542	3.019	-0.138	0.001
	h	X^1	Inverted	0.182	0.002	373.47	0.0002		
9	*h*	$X^{0.6}$	Cubic	0.236	0.018	573.105	175.163	-7.5168	0.0376

From the results displayed in Table 4-13, we can see that the best fitted regression model is different for different pairs of correlated variables, with the cubic model being the winner in most cases, followed by the exponential model. However, in all cases, R^2 is very low, which means that the regression curve explains very little of the variation of the data. Therefore, unfortunately, we were not able to establish a firm regression model that can describe the nature of the relationship between any pair of correlated variables.

4.9 Conclusions

The analysis of the descriptive statistics performed in Section 4.2 leads to the conjecture that on average, customers who place larger orders, order less frequently, face larger company delivery delays and respond to these delays with larger order cancellation percentages. This conjecture is also supported by the positive test for the existence of correlation between customer service and order fill rate performed in Section 4.6. This implies that multi-item inventory control models which assume order fill rates that depend on order quantities are a good representation of reality. The analysis in Section 4.2 further shows that the order interarrival times and monetary values exhibit low to moderate variability, whereas the company delivery delays and the resulting customer order cancellation percentage exhibit moderate to high variability. The elevated variability in customer service is most likely due to the highly disruptive effect of stockouts, which leads to long delivery delays. A better design of the stocking policy used by the company might help reduce some of this variability.

The trend analysis performed in Section 4.3 showed that different customers exhibit demand trends of different size and sign from year to year. This means that the ordering behavior varies from customer to customer; therefore, studies that rely on analyzing the behavior of a set of customers and then using the results of this analysis to infer the behavior of other customers should be received with caution.

The analysis of Section 4.4 showed the customer order interarrival times and monetary value can be well described by the Weibull distribution. The interarrival time distributions, in particular, are skewed to the left and deviate from the exponential distribution although not dramatically. This does not mean that inventory control models assuming exponentially distributed interarrival times are necessarily inaccurate; however, such models should certainly be used with caution when the interarrival times deviate from the exponential distribution.

The tests for the existence autocorrelations in the customer order data performed in Section 4.5 showed that the customer order interarrival times and monetary values are not auto correlated, which for all practical purposes means that they are independent. This is good news for the vast number of inventory management researchers who have assumed independently distributed demands in their models.

The tests for the existence of correlations between the customer service variables and the future customer order variables performed in Section 4.7 showed that for two out of nine customers there is no statistical evidence that stockouts affect the rate of their future orders. For the remaining seven customers, however, there is statistically significant evidence that stockouts affect either the number of days until their next order, *e*, or the monetary value of their next order, *h*, or both. More specifically, for three customers stockouts affect both *e* and *h*, for three other customers they affect only *e*, and for one customer they affect only *h*; therefore, it seems that most customers who experience stockouts extend the number of days until their next order and some lower the amount of their next order. Moreover, the tests showed that the frequency of future orders is almost equally affected by the fraction of the current and past orders that were cancelled and the maximum as well as the weighed average delivery delay of the current and past orders. The monetary value of future orders, on the other hand, is mostly affected by the fraction of the current and past orders that were cancelled. Finally, three out of the seven customers who exhibited significant correlations had a fairly long memory of the disservice associated with a stockout, whereas the remaining four customers had no memory of past stockouts as far as their future order behavior is concerned. In summary, stockouts do have a negative effect on the rate of future demand for most of the customers, but this effect differs from one customer to another. This implies that what may alleviate the problem for one customer may not work for another customer.

Finally from regression analysis performed in Section 4.8, we were not able to establish a firm regression model that can describe the nature of the relationship between any pair of correlated variables. This means that stylized mathematical models that assume a certain functional form of the dependence of the rate of future customer orders on stockouts should be received with caution, unless they are based on empirical evidence.

There are several issues that we did not take into account in this study. We did not analyze in detail the items in each order so we did not take into account whether the customers accepted item substitution in case of a stockout, whether the company offered a price discount for the out-of-stock items, whether the cancelled items of an order were purchased from another wholesaler or were included in a subsequent order, etc. We also did not directly make the distinction between an order for an expensive item and an order for many less expensive items of equal worth. Future research should be directed towards including such details in the analysis. It would also be worth while to explore why none of the regression curves that we tried models well the adverse effect of stockouts on future demand. Including more independent variables in our model, such as the price at which the items were sold, and trying out multivariable regression models might help towards this direction.

Chapter 5 Thesis Summary

This thesis was motivated by the need to quantify the indirect cost of a stockout, which is related to the loss of customer goodwill following a stockout. Such a loss may lead to a decline in future demand and market share of a firm, especially in a competitive market environment. We examined the issue of quantifying the indirect costs of stockouts from three different perspectives covered in Chapters 2-4, respectively.

In Chapter 2, we revisited the classical Economic Order Quantity model with backorders that are being penalized with a backorder penalty cost coefficient, b. For this model, which we referred to as the PB (penalized backorders) model, we proposed to infer the value of b by connecting b to the loss in the long-run average demand rate which is affected by backorders according to Schwartz's alternative PD (perturbed demand) model. We applied this procedure to the classical PB model and three variations of it in which we replaced the explicit fixed order cost with a constraint on the order quantity, the interorder time, and the starting inventory in each cycle, respectively. We found that for the classical PB model and the variation of the PB model with the minimum starting inventory in each cycle, the optimal fill rate is always either one or zero, which implies that the inferred backorder penalty cost is either infinite or zero, respectively. In the first case, the optimal order quantity is finite, whereas in the latter case it is infinite. For the other two variations, the optimal fill rate is always either one or a finite number between zero and one, which implies that the inferred backorder penalty cost is either infinite or a positive finite number which depends on the model parameters, respectively. In both cases, the optimal order quantity is finite.

Future research following this work could be directed toward repeating this procedure for other PD models, for example models that assume that the long-run average demand rate is either a different function of the long-run average fill rate than the one given by equation (2.2), or a function of some other customer service related performance measure, such as the long-run average backorder waiting time or number of backorders. In any such model the parameters of the perturbed demand function would have to be estimated. As was mentioned in Section 2.2, Schwartz (1966) proposed a procedure for measuring parameters A and B in his model from observed demand data. In general, this is not an easy task; however, it is a better defined task that picking a value for b. Of course, a broader question is, which perturbed demand model is correct? To answer this question, one would have to try different models and use statistical analysis of real demand data to identify the most appropriate model. We did this in Section 4.8 but we were not able to establish a firm regression model that can describe the nature of the relationship between stockouts and future demand. Finally, two other worthwhile directions for future research following this work would be to include direct backorder costs besides the indirect loss-of-customer-goodwill costs, to examine models with lost sales instead of order backlogging, and to extend this analysis to stochastic inventory models.

In Chapter 3, we analyzed a discrete-time infinite horizon inventory model in which two suppliers compete for a single customer on product availability. We formulated the problem of finding optimal stationary ordering policies for both suppliers at equilibrium as a stochastic dynamic game, and we numerically solved the resulting optimality conditions for several instances of this problem. The results indicate that both suppliers must follow the same type of policy, which can be characterized as an order-up-to policy. The order-up-to levels are generally different for each supplier and as the numerical examples suggested, they are independent of the competitor's inventory position.

It may not be that difficult to prove that the optimal ordering policy for each supplier, given that his competitor uses a stationary ordering policy, is an order-up-to policy. What is certainly much more demanding to prove is the existence and uniqueness of a NE, as the analysis of the simple two-credibility-level system in Section 3.4 showed.

The proposed inventory model could be extended and modified in a number of ways. Firstly, we could consider the situation where there is no information sharing between the two suppliers. In this case each supplier must decide his replenishment orders without knowing the inventory or service level of his competitor. Another meaningful extension of our model would be to set the selling values as decision variables. In this way the customer's next supplier choice would be affected not only by previous service levels, but by pricing policies as well. Finally, we could generalize the model for more than two suppliers.

Finally, in Chapter 4, we sought empirical evidence that stockouts do adversely affect future by performing a thorough statistical analysis of historical customer order and delivery data of a tool wholesaler and distributor over a period of four years. The tests for the existence of correlations between the customer service variables and the future customer order variables showed that for two out of nine customers there is no statistical evidence that stockouts affect the rate of their future orders. For the remaining seven customers, however, there is statistically significant evidence that stockouts negatively affect the number of days until their next order and/or the monetary value of their next order. Moreover, the tests showed that the frequency of future orders is almost equally affected by the fraction of the current and past orders that were cancelled and the maximum as well as the weighed average delivery delay of the current and past orders. The monetary value of future orders, on the other hand, is mostly affected by the fraction of the current and past orders that were cancelled. Finally, three out of the seven customers who exhibited significant correlations had a fairly long memory of the disservice associated with a stockout, whereas the remaining four customers had no memory of past stockouts as far as their future order behaviour is concerned. In summary, stockouts do have a negative effect on the rate of future demand for most of the customers, but this effect differs from one customer to another. This implies that what may alleviate the problem for one customer may not work for another customer. Unfortunately we were not able to establish a firm regression model that can describe the nature of the relationship between any pair of correlated variables. This means that stylized mathematical models that assume a certain functional form of the dependence of the rate of future customer orders on stockouts should be received with caution, unless they are based on empirical evidence.

There are several issues that we did not take into account in Chapter 4. We did not analyze in detail the items in each order so we did not take into account whether the customers accepted item substitution in case of a stockout, whether the company offered a price discount for the out-of-stock items, whether the cancelled items of an order were purchased from another wholesaler or were included in a subsequent order, etc. We also did not directly make the distinction between an order for an expensive item and an order for many less expensive items of equal worth. Future research should be directed towards including such details in the analysis. It would also be worth while to explore why none of the regression curves that we tried models well the adverse effect of stockouts on future demand. Including more independent variables in our model, such as the price at which the items were sold, and trying out multivariable regression models might help towards this direction.

Bibliography

Anderson, E. T., G. J. Fitzsimons, D. I. Simester. 2006. Measuring and mitigating the cost of stockouts. *Management Science* 52(11) 1751-1763.

Argon, N. T., R. Gullu, N. Erkip. 2001. Analysis of an inventory system under backorder correlated deterministic demand and geometric supply process. *International Journal of Production Economics* 71 247-254.

Badinelli, R. D. 1986. Optimal safety-stock investment through subjective evaluation of stockout costs. *Decision Sciences* 17 312-328.

Bernstein, F., A. Federgruen. 2004a. A general equilibrium model for industries with price and service competition. *Operations Research* 52(6) 868-886.

Bernstein, F., A. Federgruen. 2004b. Dynamic inventory and pricing models for competing retailers. *Naval Research Logistics* 51(2) 258-274.

Cachon, G.P., S. Netessine. 2005. Game theory in supply chain analysis. *TutORials in Operations Research*, Paul Gray, Series Editor, INFORMS, New Orleans, 200-233.

Caine, G. J., R. H. Plaut. 1976. Optimal inventory policy when stockouts alter demand. *Naval Research Logistics Quarterly* 23 1-13.

Campo, K., E. Gijsbrechts, P. Nisol. 2000. Towards understanding consumer response to stock-outs. *Journal of Retailing* 76(2) 219-262.

Campo, K., E. Gijsbrechts, P. Nisol. 2003. The impact of retailer stockouts on whether, how much and what to buy. *International Journal of Research in Marketing* 20 273-286.

Campo, K., E. Gijsbrechts, P. Nisol. 2004. Dynamics in consumer response to product unavailability: Do stock-out reactions signal response to permanent assortment reductions? *Journal of Business Research* 57 834-843.

Carmon, Z., J. Shanthikumar, T. Carmon. 1995. A psychological perspective on service segmentation models: The significance of accounting for consumers' perceptions of waiting and service. *Management Science* 41 1806-1815.

Chang, Y. S., and P. Niland. 1967. A Model for Measuring Stock Depletion Costs. *Operations Research* 15 427-447.

Charlton, K., A. S. C. Ehrenberg. 1976. An experiment in brand choice. *Journal of Marketing Research* 8 152-160.

Dana, J.D. Jr., N.C. Petruzzi. 2001. Note: The newsvendor model with endogenous demand. *Management Science* 47 (11) 1488-1497.

Dion, P. A., L. M. Hasey, P. C. Dorin, J. Lundin. 1991. Consequences of inventory stockouts. *Industrial Marketing Management* 20 23-27.

Dion, P. A., P. M. Banting. 1995. Buyer reactions to product stockouts in business to business markets. *Industrial Marketing Management* 24 341-350.

Emmelhainz, M. A., J. R. Stock, L. W. Emmelhainz. 1991. Consumer responses to retail stockouts. *Journal of Retailing* 67 2 138-147.

Ernst, R., M. Cohen. 1992. Coordination alternatives in a manufacturer/dealer inventory system under stochastic demand. *Production Operations Management* 1 254-268.

Ernst, R., S. G. Powell. 1995. Optimal inventory policies under service-sensitive demand. *European Journal of Operational Research* 87 316-327.

Ernst, R., S. Powell. 1998. Manufacturer incentives to improve retail service levels. *European Journal of Operational Research* 104 437-450.

Fitzsimons, G. J. 2000. Consumer response to stockouts. *Journal of Consumer Research* 27(2) 249-266.

Gans, N. 2002. Customer loyalty and supplier quality competition. *Management Science* 48(2) 207-221.

Gardner, E. S. 1980. Inventory theory and the Gods of Olympus. *Interfaces* 10 42-45.

Gaur, V., Y.-H. Park. 2007. Asymmetric consumer learning and inventory competition. *Management Science* 53(2) 227-240.

Hall, J., E. Porteus. 2000. Consumer service competition in capacitated systems. *Manufacturing and Service Operations Management* 2 (2) 144-165.

Hanssmann, F. 1959. Optimal inventory location and control in production and distribution networks. *Operations Research* 7 483-498.

Hill, T. W., Jr. 1976. A simple perturbed demand inventory model with ordering cost. *Management Science* 23(1) 38-42.

Hines, W. W., D. C. Montgomery. 1990. *Probability and Statistics in Engineering and Management Science*. Wiley, New York, NY.

Hopp, W. J., M. L. Spearman. 2000. Factory *Physics: Foundations of Manufacturing Management*. McGraw-Hill, Boston, MA.

Hui, M., D. Tse. 1996. What to tell customer in waits of different lengths: an integrative model of service evaluation. *Journal of Marketing* 60 81-90.

Kumar, P., M. Kalwani, M. Dada. 1997. The impact of waiting time guarantees on customers' waiting experiences. *Marketing Science* 16 295-314.

Li, L., 1992. The role of inventory in delivery time competition. *Management Science* 38(2) 182-197.

Liu, L., Shang, W., Wu, S. 2007. Dynamic competitive newsvendors with service-sensitive demands. *Manufacturing & Service Operations Management* 9 84-93.

Lippman, S.A., K.F. McCardle. 1997. The competitive newsboy. *Operations Research* 45(1) 54-65.

Motes, W. H., S. B. Castleberry. 1985. A longitudinal field test of stockout effects on multi-brand inventories. *Journal of the Academy of Marketing Science* 13 54-68.

Munichor, N., A. Rafaeli. 2005. Numbers or apologies? Customer reactions to tele-waiting time fillers. *Journal of Applied Psychology* (in press).

Netessine, S., N. Rudi, Y. Wang. 2006. Inventory competition and incentives to backorder. *IIE Transactions* 38 883-902.

Nielsen, A. C. 1968a. The out-of-stock study: A crusade against stockouts. *Progressive Grocer* 47(10) 55-71.

Nielsen, A. C. 1968b. Out-of-stock study: part II – growing problems of stockouts verified by Nielsen research. *Progressive Grocer* 47(11) 49-64.

Oral, M. 1981. Multi-item inventory management with monetary objective function. *AIIE Transactions* 13(1) 41-46.

Oral, M., M. S. Salvador, A. Reisman, B. V. Dean. 1972. On the evaluation of shortage costs for inventory control of finished goods. *Management Science* 18 B-344-351.

Robinson, L. 1991. Appropriate inventory policies when service affects future demands. Working paper, Johnson Graduate School of Business, Cornell University, Ithaca, NY.

Schary, P. B., B. W. Becker. 1978. The impact of stock-out on market share: Temporal effects. *Journal of Business Logistics* 1(1) 31-43.

Schary, P. B., M. Christopher. 1979. The anatomy of a stock-out. *Journal of Retailing* 55(2) 59-70.

Schwartz, B. L. 1965. Inventory models in which stockouts influence subsequent demand. Technical Report No. 12, Graduate School of Business, Stanford University, Palo Alto, California.

Schwartz, B. L. 1966. A new approach to stockout penalties. *Management Science* 12(12) B538-B544.

Schwartz, B. L., 1970. Optimal inventory policies in perturbed demand models. *Management Science* 16(8) B509-B518.

Straughn. K. 1991. *The Relationship Between Stock-Outs and Brand Share*. Ph.D. Dissertation, Florida State University.

Taylor, S. 1994. Waiting for service: The relationship between delays and evaluations of service. *Journal of Marketing* 58 56-69.

Urban, T. L. 1995. Inventory models with demand rate dependent on stock and shortage levels. *International Journal of Production Economics* 40 21-28.

Walter, C. K., J. R. Grabner. 1975. Stockout cost models: Empirical tests in a retail situation. *Journal of Marketing* 39 56-58.

Zhou, R., D. Soman. 2003. Looking back: Exploiting the psychology of queueing and the effect of the number of people behind. *Journal of Consumer Research* 29 517-529.

Zinszer, P. H., J. A. Lesser. 1981. An empirical evaluation of the role of the role of stock-out on shopper patronage processes. *Advances in Consumer Research* 8 221-224.

Zipkin, P. H. 2000. *Foundations of Inventory Management*. McGraw-Hill, Boston, MA, 29-72.