Incorporating Buyer Behavior to Drive Supply Chain Operations

by

NIHAT ALTINTAS

Submitted to the Tepper School of Business

in Partial Fulfillment of the Requirements for the Degree of

DOCTOR OF PHILOSOPHY

at the

CARNEGIE MELLON UNIVERSITY

April 2006

Thesis Committee

Professor Sridhar Tayur (co-chair)
Professor Michael Trick (co-chair)
Professor Sunder Kekre
Professor Alan Montgomery (outside reader)

Abstract

Exogenous demand assumptions provide accurate results at the retail level. As we go in the supply chain, the orders of the buyers get more complicated. The orders are influenced by both the decisions of the supplier and the costs of the buyer's operations. Therefore, it is critical for a supplier to understand the ordering behavior of a buyer in order to manage her operations. We consider a two-stage supply chain where the buyer faces a stationary and stochastic demand from his customers. The buyer places an order to maximize his profits. We provide a theoretical and empirical framework to analyze the buyer behavior. We study two practical problems: design of quantity discounts and handling reported forecasts. Our goal is to improve the supplier's operations through a better understanding of the buyer's ordering behavior.

In the first part, we study an all-unit quantity discount problem under stochastic demand for a single-item in a single supplier, single buyer setting. First, we analyze the buyer's problem. For a single period, we derive the buyer's optimal policy, which we call a three-index policy. We investigate the structure of the optimal policy for the infinite horizon problem and show that higher index policies with complex order structures may turn out to be optimal. We characterize the conditions under which such order structures are observed and identify two strategies, *wait-and-see* and *buy-and-hold*, that complicate the optimal policy. With a numerical analysis, we show that for the infinite horizon problem the performance of the best three-index policy is within 4.5% of the optimal, with an average deviation of no more than 1%. Next, we look into the supplier's problem. We suggest discount schemes that the supplier may use to moderate the buyer's ordering behavior under different transportation costs. We investigate the effect of increasing transportation costs on the discount schemes offered by the supplier. We also discuss potential limitations of quantity discounts and provide pitfalls in designing quantity discounts.

In the second part, we provide a data mining analysis of forecasting patterns of multiple buyers (auto manufacturers) from a large auto parts supplier. As part of collaborative forecasting, forecasts of future orders are used as inputs for a series of decisions. We define the complexities that are captured from our data set. We develop the daily flow analysis to obtain accuracy ratios of forecasts as a performance measure for buyers. We also demonstrate the application of some recent developments in clustering and pattern recognition analysis which can have a significant impact on the performance analysis of buyers. In our empirical analysis, we show that buyers are consistent with their forecast behavior over time. Some buyers consistently over or underestimate their orders.

In the third part, we look at the factors behind the poor forecast performance of the buyers. Unfortunately buyer forecasts cannot be used directly as we show in the second part. They may be biased since the buyer may want to mislead the supplier into believing that orders may be larger than expected to secure favorable terms or simply because the buyer is a poor forecaster. There are several unusual elements in our problem: Analysts typically observe the actual process which may be biased due to asymmetric loss function. Also units are discrete not continuous. We believe the forecasting process can be modeled in a multi stage process. The buyer first computes the distribution of demand at a future date which we assume to follow an ARMA(1,1) model. The buyer then submits an integral forecast (multiples of lot sizes), which minimizes his expected loss due to forecast errors. From data we show that the buyer may bias his forecasts under an asymmetric loss functions. We provide an estimator which can be used by the supplier to estimate the forecast generation model of the buyer by looking at his forecasts.

Acknowledgements

I would like to first thank my parents, Hediye and Ali Altintas for their constant love and support. They have always been there for me. My parents always wanted me to have a doctoral degree. Without their support, I would not have started this journey. Thanks to my brother who is always my best friend and introduced us to lovely Ilknur and to my nephew Mert. I am planning to spend much more time with my two year-old nephew. I have to make sure that he does not remember me as a boring uncle who always works late at night.

I wish to express my gratitude to my advisors for their patience, guidance and support over my tenure at Carnegie Mellon University. I would like to thank Professor Sridhar Tayur for sharing his expertise and wisdom and helping me make the right decisions about my career. I am thankful for all his efforts to motivate me and keep me on track. I would like to thank Professor Michael Trick for putting me into the field of data mining which is one of the best things that happened to me at Carnegie Mellon University. I would have never known I might be a fan of Data Mining without him. I want to thank Professor Alan Montgomery who has shaped my research for the last two years. Without his support, it would be impossible to learn so much about data analysis.

I want to also thank Professor Sunder Kekre for his continuous guidance and support. I was lucky to be a teaching assistant for him. It was an interesting experience to design and update the Operations Management course with the most recent material each year.

I want to thank Elif who made Pittsburgh a better place to live during the last three years. I was fortunate to meet Ensar, whom I had great time with. I want to also thank Cevdet, Borga and Sinan for making life fun and keeping spirits high. I will have great memories from my life in Pittsburgh.

Table of Contents

1 Introduction **1**

1.1 Design of Quantity Discounts 3

1.2 Understanding Forecast Behavior of the Buyers 5

1.3 Predicting Orders from Forecasts 7

2 Quantity Discounts under Demand Uncertainty **9**

2.1 Model Definition 11

2.1.1 Buyer's Problem 12

2.1.2 Supplier's Problem 14

2.2 Literature Review 15

2.3 Structural Analysis of the Buyer's Optimal Response 18

2.3.1 Optimal Policy for the Single Period Problem 18

2.3.2 Infinite Horizon Problem 25

2.4 Numerical Analysis of the Buyer's Problem 27

2.4.1 Why Can a Three-Index Policy Be Non-Optimal? 28

2.4.2 How Bad Can a Three-index Policy Perform? 32

2.5 A Closer Look at the Buyer's Orders 35

2.6 Design of Quantity Discounts 37

2.6.1 Numerical Analysis of the Supplier's Problem 39

2.6.2 Minimum Order Quantity Problem 42

2.7 Conclusion and Future Research 47

3 A Data Mining Approach to Forecast Behavior of Manufacturers in Automotive Industry **49**

3.1 Introduction 49

3.2 Environment Analysis 51

3.2.1 Automotive Suppliers 51

3.2.2 Order Forecasts . . . 52
3.2.3 Database Information . . . 53
3.2.4 Objective . . . 54
3.3 Literature Review . . . 55
3.4 Understanding Customer Forecasts . . . 56
3.4.1 Order Definition . . . 57
3.4.2 Daily Flow Analysis . . . 60
3.5 Characterizing Customer's Forecast . . . 64
3.5.1 Comparing Customer Performances . . . 65
3.5.2 Customer Clustering Analysis . . . 66
3.6 Managerial Use . . . 72
3.7 Conclusion and Future Research . . . 75

4 Using Customers' Reported Forecasts to Predict Future Sales 77
4.0.1 Literature Review . . . 80
4.1 Problem Environment . . . 81
4.1.1 Demand Model . . . 82
4.1.2 Forecast Generation Model . . . 83
4.2 Supplier's Problem . . . 88
4.3 Estimation . . . 90
4.3.1 Mechanism of the Slice Sampler . . . 90
4.3.2 Our Sampler for Estimation . . . 91
4.3.3 Example . . . 95
4.4 Hierarchical Model . . . 97
4.4.1 Example . . . 99
4.5 Conclusion and Future Research . . . 101

5 Appendix 103
5.1 Appendix for Chapter 2 . . . 104
5.1.1 Sensitivity Analysis for the Single Period Problem . . . 104
5.1.2 Analysis of Order Quantities in an Ordering Interval . . . 107
5.1.3 N-Period Problem . . . 110
5.1.4 Effect of Increasing Q on Minimum Order Quantity Problem . . . 112
5.1.5 Effect of Increasing Discount Rate on Quantity Discount Problem . 116
5.1.6 Stochastic Quantity Discounts with the Disposal Option . . . 117
5.2 Appendix for Chapter 4 . . . 126
5.2.1 The conditional distribution of X_1 . . . 126

5.2.2 Matrix Form of the problem 126

List of Figures

1.1 Orders of two different Wal-Mart stores from Heinz. 2

2.1 The cost functions of different ordering strategies and cost of the optimal policy for Poisson(5), $c_0 = 1.0$, $c_1 = 0.7$, $\bar{p} = 2.0$, $\bar{h} = 0.3$, and $Q = 10$ for various initial inventory positions. We use $(a)^+$ to denote $\max\{0, a\}$. 20

2.2 For different initial inventory positions x, order quantity $(q^*(x))$ and inventory position after an order is placed $(y(x))$ are displayed for the optimal policy with the parameter setting of Figure 2.1 (Poisson(5), $c_0 = 1.0$, $c_1 = 0.7$, $\bar{p} = 2.0$, $\bar{h} = 0.3$, and $Q = 10$). 24

2.3 Structure of the optimal policy for a range of Q and c_0 values. 33

2.4 Demand is Discrete Uniform $(a, 40-a)$, $c_0 = 1.0$, $c_1 = 0.7$, $p = 0.4$, $h = 0.15$, and $Q = 50$. Bounds are provided for 95% confidence level based on $10,000$ experiments for each a. 36

2.5 The effect of the demand variance on the order variance. Demand is Discrete Uniform with mean of 20, $c_0 = 1.0$, $c_1 = 0.7$, $p = 0.4$, $h = 0.15$, and $Q = 50$. We change the demand variance slightly while keeping the mean demand the same. Notice that OV changes dramatically and in the opposite direction. . 37

2.6 The supplier's optimal discount scheme for $C = 45$ and different K values. Demand is Discrete Normal$(30, 5)$, $c_1 = 1.0$, $p = 0.05$, and $h = 0.30$. 40

2.7 The supplier's optimal discount scheme for different p and h values. Demand is Discrete Normal$(30, 5)$ and $c_1 = 1.0$. 41

2.8 The supplier's optimal discount scheme for different σ values. Demand is Discrete Normal$(30, \sigma)$, $c_1 = 1.0$, $p = 0.05$, and $p = 0.3$. 43

2.9 Values of α for $Q > Q^* = 55$. Demand is Poisson(20), $p = 0.45$, and $h = 0.15$. 45

2.10 Minimum discount percentages $z^*(Q)$ that guarantees an (α) two-index policy. Demand is Uniform $(a, 40 - a)$, $p = 0.45$, $h = 0.15$, and $\alpha < 0.001$. . . . 46

3.1 11% of the customers accounts for 80% of the revenue 54

3.2 The dashed curve corresponds to a customer who falls into the true-estimator bin 85% of the time. The solid curve exhibits the behavior of a customer who over-estimates the size of his orders most of the time. The dotted curve represents the ordering pattern of a customer who under-estimates the size of his actual orders. 63
3.3 Average customer distributions for each cluster for K-Medoid Analysis. . . . 67
3.4 K-Medoid Analysis for years 2000 and 2001. Each point represents a customer. 68
3.5 K-Medoid Analysis for years 2000 and 2001 for top 11% of the customers. Each point represents a customer. 69
3.6 Average customer distributions for each cluster for Kohonen Network Analysis. 70
3.7 Kohonen Network Analysis for years 2000 and 2001. Each point represents a customer. 71
3.8 Kohonen Network Analysis for top 11% of the customers who represent 80% of the total revenue. Each point represents a customer quarter result. . . . 71
3.9 The average distribution for all customers 73

4.1 The forecast updates at different order dates 1-Aug, 8-Aug and 15-Aug. . . 82
4.2 Manufacture first observes demand and submit a forecast to the supplier . . 83
4.3 The optimal k-step ahead forecasts at time t for $\mu = 12$, $\sigma = 8$, $\phi = 0.7$, $\theta = 0.1$, $X_t = 6$, $\epsilon_t = 3$. The solid line represents an unbiased forecast with $c_u = c_o = 1$. 86
4.4 The forecast vector $\hat{X}$ and the posterior distributions for μ, ϕ, θ, σ and c_u . 96
4.5 The actual and corrected forecast after the analysis at $t = 1$. 97
4.6 The mean vector $\hat{\Lambda}$ for the hierarchical model with 16 customers and a single part 100

5.1 For Discrete Uniform$[16, 20]$, with $c_0 = 1$, $c_1 = 0.7$, $p = 0.4$, $h = 0.15$, $\bar{p} = 2$, $\bar{h} = 0.3$, and $Q = 20$, the first-period optimal policy for the two-period problem. 111
5.2 The cost functions of different ordering strategies and cost of the optimal policy for Poisson(6), $s_d = 0$, $c_0 = 1.0$, $c_1 = 0.5$, $\bar{p} = 1.5$, $\bar{h} = 0.4$, and $Q = 20$, for various initial inventory positions. 121

List of Tables

2.1 A sample of literature that studies the quantity discount problem. Unless otherwise stated, papers study all-unit discounts under a deterministic demand setting. 16
2.2 Performance of the best three-index* policies out of all combinations of $c_0 \in \{1, 1.3, 1.6\}$, $c_1 = 0.7$, $p \in \{0.15, 0.3, 0.45, 0.6, 0.75\}$, $h \in \{0.15, 0.3, 0.45, 0.6, 0.75\}$, and $Q \in \{10, 40, 70, 100\}$. 34

Chapter 1

Introduction

Understanding customer demand has always been a challenge in supply chain management. There are different aspects of orders placed at each level in a supply chain. It is common and perhaps sufficiently accurate to model end item level demand at the buyer as an exogenous demand distribution. As we go “up” in the supply chain, the traditional models and exogenous demand assumptions are often grossly violated. Much of the operations literature still makes exogenous demand assumptions at the supplier level. However, this turns out to be an unrealistic assumption when we look at real order data from the suppliers.

During my doctoral studies at Carnegie Mellon University, I have participated in industry projects with two major suppliers: Bosch and Heinz. We analyze buyers’ ordering behavior to improve the supplier’s supply chain operations. Figure 1 shows orders that are placed by two different Wal-Mart stores at Heinz. As we can see from Figure 1, there is heterogeneity in the customer orders between different buyers and the orders do not follow a well-behaved demand distribution. Therefore, assuming exogenous demand at the supplier can be quite misleading.

The orders of the buyer are influenced by several factors other than the end-item level demand. The buyer can have different costs associated for each action and there might be different pricing and penalty schemes offered by the supplier. In this case, a buyer is

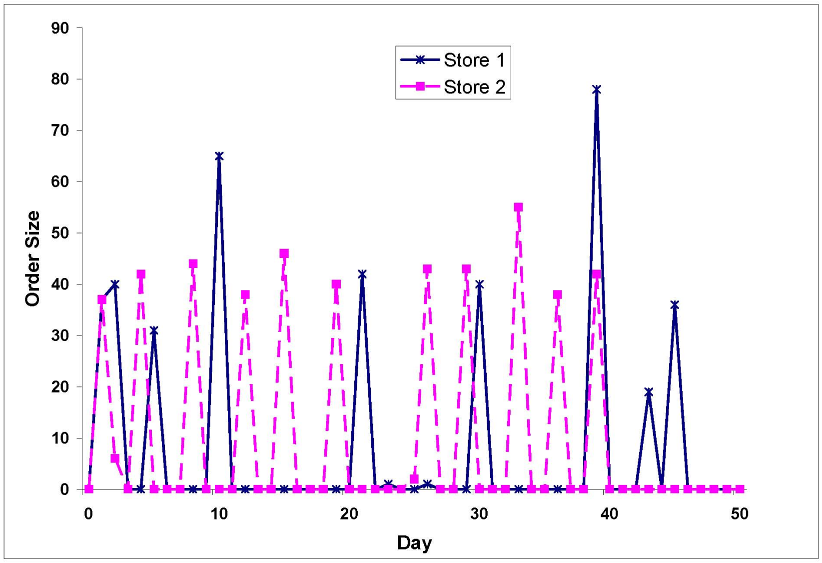


Figure 1.1: Orders of two different Wal-Mart stores from Heinz.

expected to have some level of decision process to maximize his profits. The management of operations at the supplier level requires a better understanding of the decision process of the buyer. Therefore, a realistic model should assume that the orders generated by a buyer are a result of some rational decision process. In this thesis, we study improving supplier's operations through better understanding of her buyer's orders. We do so by considering a two-stage supply chain where the buyer faces a stationary and stochastic demand from his customers at the end-item level. In our model, the buyer uses a decision support mechanism to generate his orders. We also provide empirical support for this rationality assumption from the real data at the supplier level.

In this thesis, we study improving supplier's operations through better understanding of her buyer's orders. We concentrate on three important topics that are motivated from the industry projects for the suppliers:

1. Design of quantity discounts

2. Understanding forecast behavior of buyers

3. Predicting orders from forecasts

In Chapter 2, we first study effective design of quantity discounts from a supplier's perspective. In Chapter 3, we analyze the forecasts of buyers (auto manufacturers) in automotive industry and show how buyers are consistently overestimating and underestimating their orders. In Chapter 4, we study the reasons for the poor forecast performance of the customers and test the hypothesis that the buyer generates his forecasts to minimize his loss. In all three cases, we consider a two-stage supply chain where the buyer faces a stationary and stochastic demand from his customers at the end-item level. We do not make any exogenous demand assumption for the orders at the supplier level. The buyer places his orders to maximize his profit (or minimize loss). Therefore the orders are generated as a result of some decision process by the buyer.

1.1 Design of Quantity Discounts

In Chapter 2, we study quantity discounts under demand uncertainty. Increasing profits in the presence of discounting is one of the biggest challenges that suppliers face today. When combined with buyers' increasing purchasing power, the fierce competition forces the suppliers to offer discounts. In addition to the competitive reasons, there are many tangible and intangible operational benefits of larger order sizes that justify the lower prices of quantity discounts that a supplier offers. The transportation costs are increasing due to increasing oil prices and capacity shortages. High fixed-cost operations such as transportation gain greatly from economies of scale. In order to decrease their own costs, the suppliers should convince their customers to place full truckload orders. One way of doing this is passing some of the savings to their customers in the form of quantity discounts.

A well planned and well executed quantity discount scheme can actually increase the supply chain efficiency and benefit the suppliers in environments with (high) fixed costs. However, designing effective discount schedules is difficult. It is important that a supplier understands how a buyer responds to a quantity discount scheme that she offers. Only after that she can design efficient and profitable discount schemes. This observation, as well as the difficulties encountered during the analysis of replenishment strategies at grocery retailers, such as Shaw's Supermarkets in Boston (Erhun and Tayur 2003), motivated us to study the buyer's problem. Subsequently, a project with Heinz reinforced our interest at the design of quantity discounts from a supplier's perspective. We jointly study the effect of discounts on optimal buyer behavior under stochastic demand and the characteristics of effective discount schemes from a supplier's perspective.

We first study the buyer's optimal policy under quantity discounts. In a single period setting, the optimal policy of the buyer is a *three-index policy* with parameters (S_0, S_1, S_{01}), where S_0 and S_1 are order-up-to levels with and without the discount, respectively. S_{01} is the inventory position where the buyer switches from the discounted price to the original price. For the infinite horizon problem, a three-index policy is optimal in most of the parameter settings. However, a general well-behaved policy no longer exists. In order to take the advantage of the discount, the buyer may defer his orders, i.e., *wait-and-see*, or forward buy, i.e., *buy-and-hold*. These strategies complicate the optimal policy and lead to higher number of indices. We then look into the design of an efficient discount scheme from the supplier's perspective. Analytically, we can determine discount schemes where the buyer's optimal policy is better behaved. We discuss the possible limitations of such schemes.

The majority of the studies in quantity discounts literature analyze quantity discounts with deterministic demand. To the best of our knowledge, there are no results on a buyer's optimal behavior for an all-unit quantity discount problem under stochastic demand for multiple periods. One of our objectives in this paper is to close this gap in the literature. The main limitation of the existing models from a supplier's perspective is that they do not consider the impact of quantity discounts on the buyer's behavior. By offering quantity discounts, for example, in the form of trade promotions, companies are likely to create artificial spikes in their demands, which in turn would result in increased logistics costs, the purchase of extra equipment, overtime, and adjustments in work force to deal with the excess demand. To make the matters worse, these artificial high-demand periods with increased costs will be followed by periods of cannibalized future demand which would impact the future revenues and profits. Therefore, in order to study the long-run impact of quantity discounts from the supplier's perspective, we should first understand buyers' responses to quantity discounts under multi-period stochastic settings and that is what we do in the first part of Chapter 2. We then focus on effective design of discounts from a supplier's perspective under stochastic demand, an area which has been widely overlooked in the literature.

1.2 Understanding Forecast Behavior of the Buyers

In Chapter 3 of my thesis, we concentrate on the forecast behavior of customers in supply chain. With the increased availability of enterprise wide databases, companies can easily share demand information with collaborative forecasting. We provide a data mining analysis of customer forecast performances in automotive industry. Based on our observation from our data analysis we investigate the reasons behind the noise added to the forecasts during information transmission in Chapter 4.

Collaborative Planning, Forecasting, and Replenishment (CPFR) is an evolution and refinement of these concepts among the players in the supply chain. CPFR is an initiative intended to improve the relationship among all participants in the supply chain through jointly managed planning and shared information. The quality and the intensity of the information exchange demands a strong commitment to cooperation from the participating organizations. The planning and forecasting components require intensive information exchange in all levels of relationships. Unsatisfactory relationships between the parties lead to inefficient information flow in the supply chain.

Our interest in the problem started with an industry joint project with a major auto parts supplier. The supplier started collecting demand information from the customers (auto manufacturers) under a collaborative forecasting initiative. Our goal is to improve the supplier's operations through a better understanding of its customers' ordering behavior. The supplier is using the forecasts of future orders as inputs for a series of decisions. We provide a data mining analysis of customer forecasting patterns of these buyers. We explore and analyze large quantities of data to discover meaningful patterns and rules for the customers. We define the complexities that are captured from our data set, developing the daily flow analysis to obtain accuracy ratios of forecasts as a performance measure for customers. We also demonstrate the application of some recent developments in clustering and pattern recognition analysis that can have a significant impact on the performance analysis of customers.

Our goal is to define a framework for analyzing the forecast performance of the customers and to provide empirical support for different forecast behaviors. In our analysis, we explore and analyze large quantities of data to discover meaningful patterns and rules for the ordering process of customers. In Chapter 3, we provide answers to the following questions:

a. Are classical forecasting techniques applicable? If not, is it possible to clean the forecasts in order to prepare the data for a deeper estimation analysis?

b. Is it possible to provide a quantitative representation of a customer's forecast performance?

c. Can we visualize the behavior of customers through time?

d. Can customers be assigned to groups? Do these groups represent significant commonality among different forecast behaviors?

e. How can we recognize different forecasting patterns? Is it possible to automate the process of anomaly detection from the data?

f. Are customers consistent with their forecast behavior? What is the general tendency in terms of over-forecasting or under-forecasting?

In our analysis we show that customers are consistent with their forecast behavior. Some customers consistently provide poor forecast performance.

1.3 Predicting Orders from Forecasts

In Chapter 3, we show that some buyers are consistently overestimating or underestimating their forecasts. In Chapter 4, we study the reasons behind the poor forecasts performance of the buyers and provide a technique to predict the future orders of the buyers by looking at their forecasts. The bias added to the forecasts by the buyer makes the use of forecasts questionable for the supplier. We show that the supplier can still extract any available information from the forecasts.

We assume that the forecasts are generated as a result of a multi-step decision process at the buyer. Due to uncertainty in his own demand, the buyer cannot provide a true estimate for his own order in advance. In order to model the end item demand at the buyer, we use an ARMA(1,1) model. Depending on his available information, the buyer first computes internal forecasts which can be modeled as some probability distribution. The buyer can consider the trade-off between overestimation and underestimation. By using a newsboy framework, the buyer can submit a forecast which minimizes his cost. Due to the issues with lot sizes the final forecast can be converted to an integer value which is the number of lots. In order handle the cases where the demand does follow an ARMA process, we assume that there might be some additional noise in the final forecast.

Our main objective is to explain the cost and demand structure of the buyer by looking at his forecasts. In this case, the buyer can remove the bias from the forecasts and do product planning accordingly. If the supplier fails to generate unbiased forecasts from the actual forecasts of a buyer, the credibility of the forecast information decreases. We provide a sampling algorithm which can be used effectively to predict the cost and demand parameters of a buyer. The supplier can then use the results to generate unbiased forecast for the buyer's orders.

We also build a hierarchical model where we study several buyers and products at the same time to understand the overall cost and demand structure of the buyers. This also gives us chance to predict the real orders better by looking at the forecasts of other buyers. The buyers of a product can inflate their orders at the same time due to some factors in the market. Therefore, it might be practical to know the overall trend in the industry.

Chapter 2

Quantity Discounts under Demand Uncertainty

The transportation costs are growing, on average, 6% annually due to increasing fuel prices and capacity shortages (Banta 2004). The changing transportation cost structure increases the pressure on supply chain managers to eliminate transportation inefficiencies. According to a recent survey of over 60 corporate executives from a wide range of industrial companies, 98% of manufacturers said that the changing cost structures have already impacted their business and 77% say their company executives are more focused on supply chain operations because of it (Industry Directions 2005). Companies who manage their transportation network effectively gain considerable competitive advantage in the market due to lower overall costs. One way of achieving this goal is minimizing inefficient use of transportation capacity by administering full truckload shipments: truckload operations are simpler to manage and lower in costs compared to less-than-truckload shipments. This approach has a possible setback: larger orders decrease the transportation costs in expense of inventory-related costs. Especially when these two costs are incurred by different parties, such as suppliers and buyers, there is a clear incentive misalignment. For the suppliers who carry the burden of transportation costs, a common practice to align incentives is to provide their customers quantity discounts. This way the suppliers pass on some of their savings due to better capacity utilization to their customers.

Although very common in practice, quantity discounts are hard to manage. Less than 10% of suppliers in a recent study feel that they are effectively managing their discount and promotion schemes (Adesso 2005). From the supplier's perspective, poorly managed discount schedules not only fail to provide the expected returns but also create additional problems like inflating the uncertainty in the system, a phenomenon known as the bullwhip effect in the literature (Lee et al. 1997). If the supplier enforces quantity discounts without considering the uncertainty in the system, these discounts are likely to create chaotic orders and increase the cost of the supplier. From the buyer's perspective a poorly managed discount schedule increases the total costs of the system. However, the current state of the academic literature on quantity discounts in stochastic environments does not equip the suppliers and their buyers to handle these inefficiencies.

These observations, as well as the difficulties we encountered during the analysis of replenishment strategies at grocery retailers, such as Shaws Supermarkets in Boston (Erhun and Tayur 2003), motivated us to study the buyers problem. Subsequently, a project to design quantity discounts at H.J. Heinz Company, a Pittsburgh-based leading global food company, reinforced our interest in the design of quantity discount schemes from a suppliers perspective.

Heinz, who manages its own transportation network, has been affected immensely by the recent increases in transportation costs. The company observed more than a 20% increase in its transportation costs during the last couple of years. Historically Heinz provides quantity discounts to its customers: when a customer orders close to a full truck (more than 42,000 pounds), he receives a discount. However, the recent increases in transportation costs forces the company to reconsider its current quantity discount practices. Heinz currently redesigns its discount mechanism to increase the average load and decrease the transportation cost.

With this paper, we jointly study two practical problems: (1) the effects of discounts on optimal buyer behavior under stochastic demand and (2) the characteristics of effective discount schemes from a suppliers perspective. We do so by considering a single-item, two-stage quantity discount model where the buyer faces a stationary and stochastic demand from his customers. We consider an all-unit quantity discount provided by the supplier to the buyer with a single price break.

Our analysis leads to one final question: is providing a quantity discount an efficient method for the supplier to eliminate small orders? It is without doubt that a careful analysis of quantity discounts improves the operations of the supplier. However, it is also true that there are cases under which quantity discounts fail completely to decrease the cost of the supplier and are far from being efficient.

The rest of the paper is organized as follows. We introduce our model and notation in § 2.1. We discuss the literature on quantity discounts § 2.2. We then study the buyer's problem under quantity discounts and provide an optimal policy in § 2.3. In § 2.4 we provide a numerical analysis to study the buyer's optimal policy in depth. We discuss why a three-index policy can be non-optimal for the infinite horizon problem (§ 2.4.1) and study the performance of a three-index policy (§ 2.4.2) under an extensive numerical analysis. We take a closer look at the buyer's orders in § 2.5 and analyze design of discounts from a supplier's perspective in § 2.6. A summary of our results and future research are discussed in § 2.7.

2.1 Model Definition

We consider a situation in which a buyer replenishes his inventory for a single item from a supplier who has no capacity restrictions. We consider an all-unit quantity discount with a single price break Q. If the quantity ordered for the item reaches Q, then a unit price of c_1 (discounted price) is paid for each item. Otherwise, the unit price is c_0 (original price). The quantity discount scheme satisfies $c_0 > c_1 > 0$ and $Q > 0$.

The buyer's demand comes from a stationary distribution and demands in different periods are independent. We represent the cumulative distribution function of demand with $F(.)$, which is not affected by the quantity discount scheme. Lead time is assumed to be zero. Inventory position, which determines the buyer's order quantity, is reviewed and decisions are made periodically. We present results for single period, multiple period, and infinite horizon problems. The order of the events is as follows:

1. At the beginning of the first period $n = 1$, the supplier announces the discount scheme (Q, c_0, c_1).

2. At each period $1 \leq n < N$: *(i)* The buyer checks his initial inventory level x_n and may place an order of q_n in order to increase his inventory level up to $y_n = x_n + q_n$. *(ii)* The supplier delivers q_n and charges the appropriate price to the buyer. The buyer first satisfies the backorders from the previous period. *(iii)* The buyer observes demand u_n. The buyer incurs the holding cost (h) for the excess inventory. If there is unmet customer demand, the buyer incurs a penalty cost (p). The inventory level for the next period is $x_{n+1} = x_n + q_n - u_n$.

3. At period N, we implicitly take into account a replenishment of the backordered demand from an alternative source with a higher price. There exists a disposal cost for the goods that cannot be sold at the end of the last period. We modify the penalty and holding costs of the last period with the cost of replenishment from the alternative source and disposal cost, respectively. We call these updated values the terminal costs. In order to eliminate trivial cases, we assume that terminal penalty cost is greater than the original price; i.e., the buyer has an incentive to eliminate all his backorders at the beginning of the last period of the planning horizon.

2.1.1 Buyer's Problem

The buyer minimizes his average operating cost, which includes procurement, inventory holding, and penalty costs. As the supplier processes and ships the order, the buyer does not incur any fixed ordering costs. At each period, the buyer decides how much to order in that particular period. While doing that, she has to balance inventory-related costs (i.e., inventory holding and and penalty costs) with the discount opportunity. For any period n, if the buyer orders up to y_n with a unit price of c_j, $j = 0, 1$, her total cost $L_{jn}(x_n, y_n)$ is

$$L_{jn}(x_n, y_n) = c_j(y_n - x_n) + H_n(y_n), \tag{2.1}$$

where $H_n(y_n)$ is the sum of all costs in the remaining periods plus the holding and penalty costs in the current period. Note that the buyer can order with the discounted price only if $q_n = y_n - x_n \geq Q$. For $n = 1, \cdots, N-1$

$$H_n(y_n) = p\int_{y_n}^{\infty}(u - y_n)dF(u_n) + h\int_0^{y_n}(y_n - u_n)dF(u_n) + E[\bar{L}_{n+1}(y_n - u_n)], \tag{2.2}$$

where $E[\bar{L}_{n+1}(y_n - u_n)] = \int_0^{\infty} \bar{L}_{n+1}(y_n - u_n)dF(u_n)$ and N is the problem horizon. The first two components of $H_n(y_n)$ are the expected inventory holding cost and penalty cost for the period n. We represent $\bar{L}_{n+1}(x_{n+1})$ as the optimal cost-to-go function for period $n+1$ on forward given state x_{n+1}.

For $n = 1, ..., N-1$, $\bar{L}_n(x_n)$ must satisfy the Bellman equation:

$$\begin{aligned}\bar{L}_n(x_n) = \min_{d_n \in \{0,1\}, y_n \geq x_n + d_n Q} & \{(c_0 + (c_1 - c_0)d_n)(y_n - x_n) \\ & + p\int_{y_n}^{\infty}(u_n - y_n)dF(u_n) + h\int_0^{y_n}(y_n - u_n)dF(u_n) \\ & + E[\bar{L}_n(y_n) - u_n]\}.\end{aligned}$$

In the last period N, we use the terminal costs $\bar{p}$ and $\bar{h}$:

$$\begin{aligned}\bar{L}_N(x_N) = \min_{d_N \in \{0,1\}, y_N \geq x_N + d_N Q} & \{(c_0 + (c_1 - c_0)d_N)(y_N - x_N) \\ & + \bar{p}\int_{y_N}^{\infty}(u_N - y_N)dF(u_N) + \bar{h}\int_0^{y_N}(y_N - u_N)dF(u_N).\end{aligned}$$

For any given x_n, the only decision variable is y_n. Let $x = (x_1, ..., x_N)$ and $y = (y_1, ..., y_N)$. Given an initial inventory position x_1, the buyer's problem can be formulated as follows:

$$\min_{y \geq x} \frac{1}{N} \bar{L}_1(x_1). \tag{2.3}$$

For the infinite horizon problem, we are using policy iteration for stochastic dynamic programming recursion to derive the optimal policy (Bertsekas 1995).

2.1.2 Supplier's Problem

The supplier pays for the fixed cost, which can be interpreted as the trucking and order processing costs of each order placed by the buyer. We assume that the supplier has unlimited capacity, and inventory holding and penalty costs are not a significant part of her costs.

Each truck has a capacity C and any number of trucks can be sent in a given period. Therefore, supplier ships $\lceil \frac{q_n}{C} \rceil$ trucks in period n, by incurring a cost of K per truck. The total fixed cost of the supplier can then be calculated as $K \sum_{n=1}^{N} \lceil \frac{q_n}{C} \rceil$. The supplier uses quantity discounts in order to decrease the cost of transportation. In period n, the supplier loses $q_n(c_0 - c_1)$ in terms of revenue if the buyer chooses to order with discount. Hence the supplier's profit is

$$B(Q, c_0, c_1) = \sum_{n=1}^{N} \left(q_n c_o - K \left\lceil \frac{q_n}{C} \right\rceil - q_n(c_0 - c_1) I_{\{q_n \geq Q\}} \right). \tag{2.4}$$

Therefore, the trade-off for the supplier is between the fixed cost and the discount that is provided to the buyer. The decision for the supplier is to find the discount scheme $(Q, \Delta_c = c_0 - c_1)$, which maximizes her average profit:

$$\max_{Q \geq 0, \Delta_c \geq 0} \frac{1}{N} B(Q, c_0, c_1). \tag{2.5}$$

2.2 Literature Review

There is vast literature related to the quantity discount problem. In Table 2.1, we provide a short summary. We refer readers to Benton and Park (1996) and Munson and Rosenblatt (1998) for extensive reviews, and to Dolan (1987) for a detailed survey of different variants of the problem from a marketing research standpoint. From a historical perspective, the interest in quantity discount problem started with the research of Buchanen (1953) and Garbor (1955), where the authors discuss the motivations for the quantity discounts. Porteus (1971) studies an incremental discount problem with a concave increasing cost function. He shows that a generalized (s, S) policy is optimal in a finite horizon problem under certain conditions, including one where the probability densities of demand in each period are Polyá densities. Sethi (1984) and Jucker and Rosenblatt (1985) provide solutions from the buyer's perspective. With a model similar to ours, Jucker and Rosenblatt study a single period problem with many discount breaks and develop an algorithm to calculate the optimal order quantities. Sethi models disposal options as a nonlinear pricing scheme under deterministic demand. He provides a method of obtaining optimal lot sizes for an entire range of disposal costs. Monahan (1984), Lal and Staelin (1984), and Lee and Rosenblatt (1986) study the economic implications for the supplier and focus on deriving pricing schemes that maximize the supplier's profit.

Another stream of quantity discount research emphasizes channel coordination. Jeuland and Shugan (1983) show that profit sharing mechanisms with quantity discounts can coordinate the supply chain. Weng (1995) and Chen, Federgruen, and Zheng (2001) show that centralized channelwide profits can be achieved in a decentralized system by different quantity discount schemes. Corbett and de Groote (2000) consider coordinating the supply chain when the buyer has some private information.

Despite their widespread use in practice, research on discount schedules under uncertainty is still in its infancy. The majority of the studies in the literature analyze quantity

Table 2.1: A sample of literature that studies the quantity discount problem. Unless otherwise stated, papers study all-unit discounts under a deterministic demand setting.

Early Results	**Buyer's Problem**	**Supplier's Problem**	**Channel Coordination**	**Survey Papers**
Buchanen (1953), Garbor (1955)	Porteus (1971)*, Sethi (1984), Jucker and Rosenblatt (1985)†	Monahan (1984), Lal and Staelin (1984), Lee and Rosenblatt (1986)	Jeuland and Shugan (1983), Weng (1995), Chen et al. (2001), Corbett and de Groote (2000)	Munson and Rosenblatt (1998), Benton and Park (1996), Dolan (1987)

* (†): Incremental (all-unit) discounts under stochastic demand

discounts with deterministic demand. One of the exceptions is Jucker and Rosenblatt (1985). The authors study all-unit discounts under stochastic demand for a single period model. In our analysis, we provide an optimal policy for a single break version of their problem. Such a policy acts as a building block for the optimal policy of the rather complicated multi-period problem. To the best of our knowledge, there are no results on a buyer's optimal behavior for an all-unit quantity discount problem under stochastic demand for multiple periods. One of our objectives in this paper is to close this gap in the literature. The main limitation of the existing models from a supplier's perspective is that they do not consider the impact of quantity discounts on the long-run profitability of the firm. By offering quantity discounts, companies are likely to create artificial spikes in their demands, which in turn would result in increased logistics costs, the purchase of extra equipment, overtime, and adjustments in work force to deal with the excess demand. To make matters worse, these artificial high-demand periods with increased costs will be followed by periods of cannibalized future demand which would impact the future revenues and profits. Therefore, in order to study the long-run impact of quantity discounts from the supplier's perspective, we should first understand the buyer's response to quantity discounts under multi-period stochastic settings and that is what we do in the first part of our paper.

We also focus on effective design of discount schemes from a supplier's perspective under stochastic demand, an area which has been widely overlooked in the literature. Cheung and Lee (2002) model shipment coordination in order to have full truckload shipments. In their analysis, when the total decrease in the inventory positions of the retailers reaches Q, the supplier restores their inventory positions back to their base-stock levels by shipping Q units in total. Cachon (1999) studies scheduled ordering policies as a way to decrease supply chain demand variability in a model with one supplier and N retailers that face stochastic demand. The retailers in his model order at fixed intervals and their order quantities are equal to some multiple of a fixed batch size.

Another stream of research related to the supplier's problem is minimum purchase commitment contracts. We refer readers to Anupindi and Bassok (1998) for a summary of results for quantity commitment contracts, and Tsay, Nahmias, and Agrawal (1998) for a review of supply chain contracts. Anupindi and Akella (1993) consider a periodic review finite horizon model in which the buyer commits to purchase of at least Q units in each period. Additional units can be purchased for a higher price but may not be delivered immediately. The authors show that the optimal policy is an S-type policy: that is, if on-hand inventory plus Q is less that S, then the inventory is adjusted up to S. Moinzadeh and Nahmias (2000) study a similar problem, but with a two-part tariff for adjustments and over an infinite rather than a finite horizon. They discuss (but do not formally prove) that a (s, S) type policy is optimal: if the inventory on hand before delivery is less than s, then the inventory is adjusted up to S. Bassok and Anupindi (1997) study a total minimum purchase commitment where the buyer commits to purchase at least K_N units over a planning horizon of N periods. In all three models, discounts increase with the committed quantity.

In our analysis, unlike Cheung and Lee (2002) and Cachon (1999), we are not making any simplified policy assumptions for the buyer's ordering behavior. Different from the second stream (Anupindi and Akella 1993, Moinzadeh and Nahmias 2000, Bassok and Anupindi

1997) the buyer does not make any commitments on the quantities to be purchased; he simply follows his optimal strategy. We show that the supplier can design quantity discount schemes to maximize her profits. Therefore, both the supplier and the buyer benefit from the discount without any initial commitment on the quantities to be purchased. The supplier has an indirect control over the buyer's order quantity by adjusting the discount scheme.

2.3 Structural Analysis of the Buyer's Optimal Response

We first analyze the buyer's optimal response to all-unit quantity discounts. We present our solution methodology for the single period problem in § 2.3.1 and discuss the infinite horizon problem in § 2.3.2.

2.3.1 Optimal Policy for the Single Period Problem

In this section, we define the single period problem and present the expected cost minimizing solution[1]. The buyer chooses an order quantity before realizing demand. Since it is a single period problem, there is no need to use subscript n:

$$L_j(x,y) = c_j(y-x) + H(y) \qquad j = 0, 1. \tag{2.6}$$

The penalty and holding cost function is

$$H(y) = \bar{p}\int_y^{\infty}(u-y)dF(u) + \bar{h}\int_0^y (y-u)dF(u), \tag{2.7}$$

which is convex. Hence, there are unique S_0 and S_1 that minimize the functions $L_0(x,y)$ and $L_1(x,y)$, respectively, such that $S_0 \leq S_1$. In order to calculate S_0 and S_1, we take the derivatives of $L_0(x,y)$ and $L_1(x,y)$ with respect to y and we achieve the newsvendor

[1] Our single period model is a special case of Jucker and Rosenblatt (1985)'s as we analyze a model with only one discount break. However, Jucker and Rosenblatt do not try to come up with a policy and work on an algorithm to calculate the optimal order quantities instead. Therefore, the policy that we introduce in this section is original.

solution:

$$S_j = F^{-1}\left(\frac{\bar{p} - c_j}{\bar{h} + \bar{p}}\right) \qquad j = 0, 1. \tag{2.8}$$

Figure 2.1 displays the cost functions of different ordering strategies for a given parameter set[2]. As can be seen from the figure, the optimal policy is a function of the initial inventory position. In this example, the buyer orders more than Q units when his initial inventory position is less than -5. By ordering up to 5, he gets the discount and avoids backorders in the selling period. The order-up-to levels are the same for all initial inventory positions less than -5. We call this order-up-to level $S_1 = 5$, which can also be derived from Equation (2.8). In order to get the discount, the buyer should order at least $Q = 10$. Therefore, $S_1 = 5$ is a feasible order-up-to level when the initial inventory position is no greater than $S_1 - Q = -5$. When the initial inventory position is -4, the buyer faces a trade-off between using the discount opportunity by ordering exactly Q units versus incurring less holding cost by ordering without the discount. He prefers to get the discount by ordering exactly $Q = 10$ units for initial inventory positions between -4 and -2. When the initial inventory is -1, the marginal return from the discount can no longer compensate the higher holding and disposal cost, and the buyer stops ordering with the discount. We call the maximum inventory position where the buyer still orders with the discounted price c_1, S_{01}. In this example, $S_{01} = -2$. For inventory positions greater than S_{01} and less than 4, the buyer orders up to $S_0 = 4$, which is the order-up-to level for price c_0 in Equation (2.8). When the initial inventory position is larger than $S_0 = 4$, he prefers not to order. The outer envelope of the feasible cost curves at a given initial inventory position determines the optimal policy. Note that even for a single period problem, the optimal cost function is not pseudoconvex in initial inventory position and not differentiable everywhere.

Next, we prove structural properties of the cost functions, $L_0(x, y)$ and $L_1(x, y)$, and determine the optimal policy using a case-by-case analysis. Our main result is that the

[2]In our examples and numerical analysis, we use demand distributions with integral non-negative values; however, we continue to represent the costs with continuous functions.

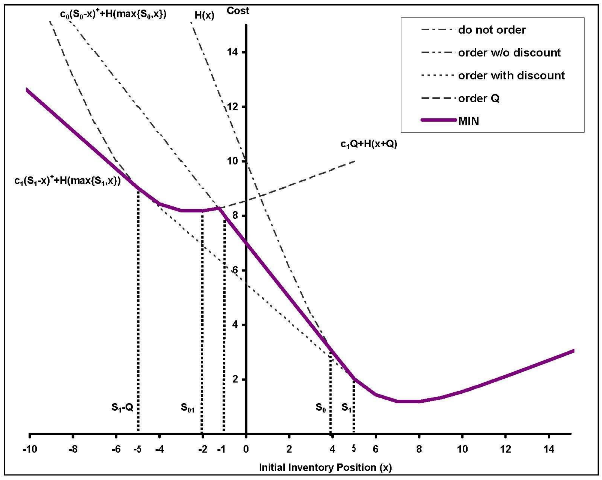


Figure 2.1: The cost functions of different ordering strategies and cost of the optimal policy for Poisson(5), $c_0 = 1.0$, $c_1 = 0.7$, $\bar{p} = 2.0$, $\bar{h} = 0.3$, and $Q = 10$ for various initial inventory positions. We use $(a)^+$ to denote $\max\{0, a\}$.

optimal policy depends on the inventory position, the discount break Q, and the minimizers of the functions $L_0(x, y)$ and $L_1(x, y)$.

Proposition 1 *For $x < S_1 - Q$, the optimal policy is to order up to S_1.*

Proof: The buyer can get the discount in two ways: he can either order up to S_1 or he can order Q. However, $L_1(x, Q) \geq L_1(x, S_1)$, since S_1 is the optimal order-up-to level for $L_1(x, y)$ and $x < S_1 - Q$. Therefore, for $x < S_1 - Q$, it is optimal to get the quantity discount by ordering an amount larger than Q. The proof analyzes two cases:

Case 1: $x < \min\{S_0, S_1 - Q\}$: In this case, the functions $L_0(x, y)$ and $L_1(x, y)$ are convex with minima at points $y = S_0$ and $y = S_1$, respectively. As $c_0 > c_1$, we have $L_0(x, S_0) > L_1(x, S_0)$. Since S_1 is the optimal order-up-to level for $L_1(x, y)$, we have $L_1(x, S_0) > L_1(x, S_1)$. Therefore, $L_0(x, S_0) > L_1(x, S_1)$.

Case 2: $S_0 \leq x < S_1 - Q$: In this case, if the buyer does not get the discount, he does not order and the total cost is $L_0(x, x)$. As $L_0(x, x) = L_1(x, x) > L_1(x, S_1)$, the optimal policy is to place an order with the quantity discount. □

Next, we consider all possible sub-cases when $x \geq S_1 - Q$. In this range, ordering with the original price may turn out to be non-optimal and the problem may reduce to a single price problem with a minimum order quantity Q.

Proposition 2 *For $x \geq S_1 - Q$, there exists a critical level S_{01}, such that:*

i. When $S_1 - Q \leq x \leq S_{01}$, it is optimal to order exactly Q units.

ii. When $x > S_{01}$, it is optimal to order without the quantity discount if necessary, i.e., the order quantity is $(S_0 - x)^+$.

Proof: We consider two cases based on the evaluation of the total cost function when the initial inventory position is equal to S_0. In the first case, when the buyer has S_0 units to start with he does not prefer to order Q units to get the discount; i.e., $H(S_0+Q) > H(S_0) - c_1 Q$. In the second case, he has a motivation to do so; i.e., $H(S_0 + Q) \leq H(S_0) - c_1 Q$.

Case 1: $H(S_0 + Q) > H(S_0) - c_1 Q$: The cost function $L_1(x, y)$ is convex and for any $y \geq S_1$ it is increasing. Hence, the buyer's optimal policy is to order exactly Q units (but no more) in case he would like to get the discount. The total cost is

$$L_1(x, x + Q) = c_1 Q + H(x + Q). \tag{2.9}$$

For $x \leq S_0$, it is optimal to order up to S_0 if the buyer does not get the discount and the total cost is

$$L_0(x, S_0) = c_0(S_0 - x) + H(S_0). \tag{2.10}$$

Ordering with the quantity discount is optimal if

$$\begin{aligned} L_1(x, x + Q) \;\leq\; L_0(x, S_0) \;&\Rightarrow\; c_1 Q + H(x + Q) \;\leq\; c_0(S_0 - x) + H(S_0) \\ &\Rightarrow\; H(x + Q) - c_0(S_0 - x) \;\leq\; H(S_0) - c_1 Q \end{aligned} \tag{2.11}$$

The right-hand side of inequality (2.11) is constant. The left-hand side, on the other hand, is convex with respect to x. For $x = S_1 - Q$, the inequality is strict as $L_1(S_1 - Q, S_1) <$

$L_1(S_1 - Q, S_0) < L_0(S_1 - Q, S_0)$. The first inequality follows from the fact that S_1 is the optimal order-up-to level for c_1, therefore ordering up to S_0 will be costlier. The second inequality follows from the fact that $c_1 < c_0$. At point $x = S_0$, $H(S_0 + Q) > H(S_0) - c_1 Q$ by assumption. Therefore, there exists a critical level S_{01} such that $S_1 - Q < S_{01} \leq S_0$ that satisfies (2.11) as an equality.

For $x > S_0$, the buyer will not order at all if he chooses not to use the discounted price. Then the total cost is

$$L_0(x, x) = H(x). \tag{2.12}$$

Ordering with the quantity discount is optimal if

$$\begin{aligned} L_1(x, x+Q) \ \leq \ L_0(x, x) \ \ \Rightarrow \ \ c_1 Q + H(x+Q) \ &\leq \ H(x) \\ \Rightarrow \ \ c_1 Q \ &\leq \ H(x) - H(x+Q). \end{aligned} \tag{2.13}$$

At $x = S_0$, $c_1 Q > H(S_0) - H(S_0 + Q)$ by assumption. Since $H(x)$ is convex and $H(x) - H(x + Q)$ is decreasing in x, for all $x > S_0$ we have $c_1 Q > H(x) - H(x + Q)$. Therefore, there does not exist any x that satisfies inequality (2.13); i.e., the buyer does not order at all when $x > S_0$.

To summarize, when $H(S_0 + Q) > H(S_0) - c_1 Q$, we have shown the following: (a.) There exists a critical level S_{01} such that $S_1 - Q \leq S_{01} \leq S_0$. (b.) When $S_1 - Q < x \leq S_{01}$, the buyer orders exactly Q units. When $x > S_{01}$, the buyer orders $(S_0 - x)^+$ units.

Case 2: $H(S_0+Q) \leq H(S_0)-c_1 Q$: For $x \geq S_0$, we use inequality (2.13). When $x = S_0$, the inequality is satisfied by the assumption of Case 2. However, as x goes to infinity, the right-hand side of the inequality becomes negative, hence $c_1 Q > H(x)-H(x+Q)$. Therefore, there exists a critical level S_{01} such that $S_{01} \in [S_0, \infty)$ that satisfies inequality (2.13) as an equality.

For $x < S_0$, from the analysis of Case 1 we know that inequality (2.11) is convex with respect to x and strict for $x = S_1 - Q$. Due to the assumption of Case 2, the inequality continues to hold at $x = S_0$. Therefore, for $x \in (S_1 - Q, S_0)$, the buyer orders exactly Q units.

To summarize when $H(S_0 + Q) \leq H(S_0) - c_1 Q$, we have shown the following: (a.) There exists a critical level S_{01} such that $S_0 \leq S_{01}$. (b.) When $S_1 - Q < x \leq S_{01}$, the buyer orders exactly Q units. When $x > S_{01}$, the buyer does not order at all. $\square$

Our first theorem combines the observations for different inventory positions and presents the optimal ordering policy for the single period problem.

Theorem 1 *(Three-index policy) The optimal order quantity ($q^*(x)$) for the single period all-unit quantity discount problem with one price break is given by the following rule:*

$$q^*(x) = \begin{cases} \max\{S_1 - x, Q\} & \textit{when } x \leq S_{01} \\ \max\{S_0 - x, 0\} & \textit{when } x > S_{01} \end{cases}$$

where $S_i = F^{-1}\left(\frac{\bar{p} - c_i}{\bar{h} + \bar{p}}\right)$. *We call this policy a* **three-index** *policy with indices* (S_0, S_1, S_{01}).

Proof: Directly follows from Propositions 1 and 2. $\square$

We have shown that it is possible to have S_{01} larger than S_0. When this is the case, the buyer exercises only the discounted price. S_0 is not utilized, and the optimal policy is a *two-index* policy with indices (S_1, S_{01}). Ordering with the original price is non-optimal for all initial inventory positions. This problem is equivalent to a *minimum order quantity problem* where an order cannot be less than Q units and there is only one price available.

As can be observed from Figure 2.2, the order quantity is monotonic in initial inventory position for single period problem. In regions A and B, the buyer orders at least Q units and gets the discount. In region C, he orders up to $S_0 = 4$. In region D, it is optimal not to order.

The optimal policy depends on the problem parameters. It is straightforward from the newsvendor formula that the order-up-to level S_i depends only on $\bar{p}$, $\bar{h}$, and c_i. However, the critical level S_{01} depends on all parameters.

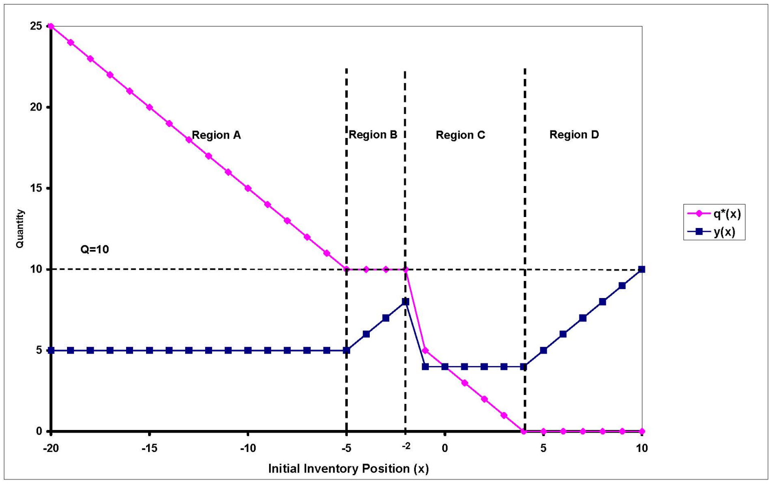


Figure 2.2: For different initial inventory positions x, order quantity ($q^*(x)$) and inventory position after an order is placed ($y(x)$) are displayed for the optimal policy with the parameter setting of Figure 2.1 (Poisson(5), $c_0 = 1.0$, $c_1 = 0.7$, $\bar{p} = 2.0$, $\bar{h} = 0.3$, and $Q = 10$).

Proposition 3 *The sensitivity results for S_{01} are as follows:*

i. The derivative of S_{01} with respect to Q satisfies $-1 < \frac{\partial S_{01}}{\partial Q} < 0$.

ii. The derivative of S_{01} with respect to c_1 while keeping c_0 constant satisfies $\frac{\partial S_{01}}{\partial c_1} < 0$.

iii. The derivative of S_{01} with respect to c_0 while keeping c_1 constant satisfies $\frac{\partial S_{01}}{\partial c_0} \geq 0$.

Proof: Available in Appendix 5.1.1. □

In the first part of the proposition, we conclude that increasing Q decreases S_{01}, with a rate less than the rate of increase in Q. So S_1 staying constant, increasing Q increases the range of the interval $(S_1 - Q, S_{01})$ where ordering Q is optimal but decreases the range where the buyer uses the discounted price, $(-\infty, S_{01})$. When Q decreases, S_{01} increases and the buyer prefers to order with the discounted price for a larger range of initial inventory positions. Therefore, the quantity discount is more attractive for the buyer when Q is smaller. In the second part of the proposition, we observe that increasing the discounted price c_1 (which means decreasing the discount rate) decreases S_{01}. When the discount rate decreases, the quantity discount becomes less attractive for the buyer. The savings through the unit cost do not balance the additional inventory risk. Therefore, he ends up using the discount opportunity for a smaller range of initial inventory positions. From part three, we observe the reverse effect for c_0. As c_0 increases (which means the discount rate increases), S_{01} increases. When the original price is considerably more expensive compared to the discounted price, the discount opportunity becomes more attractive and the buyer ends up using the discounted price for a larger range of initial inventory positions.

2.3.2 Infinite Horizon Problem

In this section, we study the structural properties of the infinite horizon problem. Our objective is to minimize the long-run average total cost. While a three-index policy continues to be optimal in many parameter settings, we observe examples where this may no longer be the case[3].

Example 1 *For Poisson*(6)*, with* $c_0 = 1.3$*,* $c_1 = 0.7$*,* $p = 0.75$*,* $h = 0.55$*, and* $Q = 30$*, the optimal policy for the infinite horizon problem is as follows:*

		initial inventory	$\leq$	*-19*	:	*order up to* **11**	*Region (I)*
-18	$\leq$	*initial inventory*	$\leq$	**-7**	:	*order* $Q = 30$	*Region (II)*
-6	$\leq$	*initial inventory*	$\leq$	*-3*	:	*order* 0	*Region (III)*
-2	$\leq$	*initial inventory*	$\leq$	*6*	:	*order up to* **6**	*Region (IV)*
7	$\leq$	*initial inventory*			:	*order* 0	*Region (V)*

Definition 1 *An* **ordering interval** *s is a minimal convex set of initial inventory positions* $[x_l^s, x_u^s]$ *at which the order quantity is larger than zero.*

When a three-index policy is optimal, there is only one ordering interval. Hence, we drop superscript s: $(x_l, x_u) = (-\infty, S_0)$ if $S_0 > S_{01}$ or $(x_l, x_u] = (-\infty, S_{01}]$ otherwise. In Example 1, there are two ordering intervals with $(x_l^1, x_u^1] = (-\infty, -7]$ and $[x_l^2, x_u^2) = [-2, 6)$. Therefore, three indices will not be sufficient to describe the optimal policy. Next we provide the definition for a k-index policy.

[3]Similar examples can be created for a finite horizon problem, with as few as two periods. Interested readers may refer to the on-line addendum, Section 5.1.3.

Definition 2 *A* **k-index policy** *is an ordering policy which cannot be specified using less than k indices.*

In Example 1, the optimal policy turns out to be a four-index policy: $(S_1^1 = 11, S_{01}^1 = -7, x_l^2 = -2, S_0^2 = 6)$. Notice that, the order quantities are *not monotonic* in initial inventory position. In region (III), it is optimal not to order, while in region (IV), the buyer starts ordering again, but this time with the original price. However, within each ordering interval, the order quantities are monotonic. Our next example shows that this does not have to be the case in general.

Example 2 *For Discrete Normal(*$30, \sigma = 3$*)*[4]*, with* $\hat{\mu} = 30$*,* $\hat{\sigma} = 3.01$*,* $c_0 = 1.0$*,* $c_1 = 0.7$*,* $p = 0.45$*,* $h = 0.15$*, and* $Q = 40$*, the optimal policy for the infinite horizon problem is as follows:*

		initial inventory	$\leq$	*-10*	:	*order up to* **30**
-9	$\leq$	*initial inventory*	$\leq$	**5**	:	*order* $Q = 40$
6	$\leq$	*initial inventory*	$\leq$	*17*	:	*order up to* **57**
18	$\leq$	*initial inventory*	$\leq$	**19**	:	*order* $Q = 40$
20	$\leq$	*initial inventory*	$\leq$	*24*	:	*order up to* **24**
25	$\leq$	*initial inventory*			:	*order* 0

In Example 2, order quantities are no longer monotonic within the ordering interval. This leads to our next definition:

Definition 3 *In an ordering interval s, a* **monotone ordering interval** *r is a minimal convex set of initial inventory positions* $[m_l^{s:r}, m_u^{s:r}]$*, in which the order quantities are larger than zero and monotonic in initial inventory position.*

[4]In order to have integral non-negative demand, we first discretize the demand function by approximating the probability of an integral demand $p_Y(y)$, where y is an integer. We then truncate this distribution and consider only the positive observations. The probability function for demand $p_Y(y)$ is equal to $\left(\frac{F(y+0.5)-F(y-0.5)}{1-F(-0.5)}\right)$. Because of discretization and truncation, the observed $\hat{\mu}$ and $\hat{\sigma}$ values differ from the original values.

An ordering interval can be divided into monotone ordering intervals. If all ordering intervals are monotone for a policy, then superscript r is dropped. In Example 1, both ordering intervals are also monotone ordering intervals. In Example 2 there is only one ordering interval with $(x_l^1, x_u^1) = (-\infty, 24)$, which consists of two monotone ordering intervals with $(m_l^{1:1}, m_u^{1:1}] = (-\infty, 5]$ and $[m_l^{1:2}, m_u^{1:2}) = [6, 24)$. The optimal policy turns out to be a five-index policy: $(S_1^{1:1} = 30, m_u^{1:1} = 5, S_1^{1:2} = 57, S_{01}^{1:2} = 19, S_0^{1:2} = 24)$.

The examples above show that a three-index policy is not necessarily optimal for the infinite horizon problem. However, we can still prove basic results that show important patterns.

Theorem 2 *In any ordering interval s, there exists at most one initial inventory position S_{01}^s, such that in an optimal policy,*

- *when $x_l^s \leq x \leq S_{01}^s$, the buyer places an order with the discount.*
- *when $S_{01}^s < x \leq x_u^s$, the buyer places an order with the original price.*

Proof: Available in the on-line addendum, Section 5.1.2. □

Theorem 2 shows that within an ordering interval, there exists a sequence from lower to higher price[5]. Note that the theorem does not make any claims about the order quantity. Within an ordering interval, the order quantities can be non-monotonic in the initial inventory position, as seen in Example 2.

Next we discuss the conditions leading to non-optimality of a three-index policy.

2.4 Numerical Analysis of the Buyer's Problem

As we have shown in Section 2.3.1, a three-index policy is optimal for the single period problem. However, this result cannot be extended to an infinite horizon. In this section, we consider different numerical settings to better understand the characteristics of the optimal

[5]We use this result in the dynamic programming solution in order to limit the search space.

policy for the infinite horizon problem. We first explore the reasons which make a three-index policy non-optimal. We then analyze the performance of the best three-index policy[6].

2.4.1 Why Can a Three-Index Policy Be Non-Optimal?

There may be several reasons to deviate from a three-index policy at optimality. In this section, we discuss two of them with numerical examples. When faced with a discount opportunity in a multi-period setting, the buyer has to consider trade-offs between the immediate and future penalty costs, holding costs, and the ability to take advantage of the discount. He can do this in two ways. He can either hold on to an order (not order or order few units) to receive discounts in a future period, i.e., **wait-and-see** strategy, or he may now order more than he actually needs and keep it for future use, i.e., **buy-and-hold** strategy. In both cases, the interactions between the periods change the dynamics of the problem.

Wait-and-See Strategy

In order to better understand the wait-and-see strategy, recall Example 1. In region (III), the buyer does not order, while in region (IV), he starts ordering again, but this time with the original price. This can be explained as follows: In region (III), the buyer cannot afford to order with the discounted price any longer due to high inventory carrying costs. Ordering with the original price is not in his best interest either. In case he orders, he would lose the discount opportunity due to two reasons. First, the discount opportunity for the items being ordered in this period would be lost. Second, the inventory position would increase and ordering with the discount would become difficult in the upcoming period. For that reason, the buyer chooses to wait until the next period in order to utilize the advantages of the quantity discount. In region (IV), however, the initial inventory position is high enough

[6]One can argue that all-unit quantity discounts have irrational characteristics: the buyer may be better off buying and disposing few units at the beginning of the period without incurring holding cost on these items. We expand our model to a setting with disposal in the on-line addendum, Section 5.1.6. Including disposal does not eliminate any of the cases that we discuss in this section.

to prevent ordering with the discounted price in the next period. Therefore, the buyer starts ordering again in order to prevent paying penalties for stock-outs. Note that region (III) divides the strategy space into two ordering intervals. In the first ordering interval, the objective is to decrease the procurement costs and in the second one it is to decrease the penalty costs. Example 3 provides additional insights about region (IV), i.e., why the buyer starts ordering again:

Example 3 *Ordering intervals and optimal order-up-to levels for Poisson(6) with* $1.2 \leq c_0 \leq 1.45$[7]*,* $c_1 = 0.7$*,* $p = 0.75$*,* $h = 0.55$*, and* $Q = 30$*. (Bold values define the optimal policy.)*

	Ordering Interval 1					**Ordering Interval 2**					**Optimal**
c_0	S_0^1	S_1^1	S_{01}^1	x_l^1	x_u^1	S_0^2	S_1^2	S_{01}^2	x_l^2	x_u^2	**Policy**
1.2	**6**	**10**	**-8**	$-\infty$	*6*	-	-	-	-	-	*Three-index*
1.25	-	**11**	**-7**	$-\infty$	*-7*	**6**	-	-	**-4**	*6*	
1.3	-	**11**	**-7**	$-\infty$	*-7*	**6**	-	-	**-2**	*6*	*Four-index*
1.35	-	**11**	**-7**	$-\infty$	*-7*	**5**	-	-	**-1**	*5*	
1.4	-	**11**	**-7**	$-\infty$	*-7*	**4**	-	-	**1**	*4*	
1.45	-	**11**	**-7**	$-\infty$	*-7*	-	-	-	-	-	*Two-index*

As c_0 increases, the buyer's behavior changes. Initially, when $c_0 = 1.2$, the discount is not high enough to tempt the buyer to wait for additional periods: he orders with the discounted price whenever he can. As the discount increases, the buyer starts to use the wait-and-see strategy; we begin to observe the second ordering interval. For moderate discounts, i.e., $1.25 \leq c_0 \leq 1.4$, the buyer still uses the original price from time to time

[7]In our proofs and examples we manipulate the original price due to a technicality. The analysis can also be performed by manipulating the discounted price and our results will remain valid. However, manipulating the original price avoids bounds on the discount ratio. Note that this is also a common practice in industry. Some suppliers fix the discounted price assuming a full truckload and provide an extra upcharge to the buyer for less than truckload orders based on the shipment size.

in order to prevent paying penalties for stock-outs. As c_0 increases further, the trade-offs change completely and the quantity discount dominates the penalty cost. It is at this point where the second order interval disappears and the buyer orders with the discounted price c_1 all the time.

Buy-and-Hold Strategy

In order to better understand the buy-and-hold strategy, recall Example 2. In order to use the discount opportunity, the buyer may order several weeks of inventory in advance. We observe two order-up-to levels for the discounted price and three-index policy is no longer optimal. When the initial inventory level is less than -10, the buyer orders up to $S_1^{1:1} = 30$. With a mean demand of 30, this can be considered as ordering weekly supply with the discounted price. When the initial inventory level is between -10 and 5, the buyer orders exactly $Q = 40$ units to get the discount. As the initial inventory level increases further, it is no longer optimal for the buyer to order Q units to get the discount. In this case, the buyer orders up to $S_1^{1:2} = 57$; i.e., the buyer procures around two weeks' worth of inventory. One of the major factors in his decision is demand uncertainty. When uncertainty is high, the future looks blurry. When uncertainty is low, he can make refined long-term decisions considering the delicate interactions between the periods. Therefore, we can anticipate that the optimal policy has higher number of indices when demand uncertainty is low.

In order to further discuss the impact of demand variability, Example 4 uses discrete normal distribution with different standard deviation values keeping the mean demand the same. When $\sigma \leq 5$ (buy-and-hold region), we observe two monotone ordering intervals. When the buyer buys and holds, he wants to have some visibility into the future. Therefore, when demand is close to deterministic, the buyer can look further in the horizon and order inventory for (around) two periods. As we further increase the standard deviation ($\sigma \geq 5.5$), the buyer no longer observes multiple order-up-to levels for the discounted price. The uncertainty in the system makes ordering arrangements for a longer period of time less attractive and the buyer orders weekly supply.

Example 4 *Monotone ordering intervals and optimal order-up-to levels for Discrete Normal(30, σ), with $c_0 = 1.0$, $c_1 = 0.7$, $p = 0.45$, $h = 0.15$, and $Q = 40$. (Bold values define the optimal policy.)*

	Observed		**Ordering Interval (O.I.) 1**											**Optimal**
			Monotone O.I. I					**Monotone O.I. II**					
σ	$\hat{\mu}$	$\hat{\sigma}$	$S_0^{1:1}$	$S_1^{1:1}$	$S_{01}^{1:1}$	$m_l^{1:1}$	$m_u^{1:1}$	$S_0^{1:2}$	$S_1^{1:2}$	$S_{01}^{1:2}$	$m_l^{1:2}$	$m_u^{1:2}$	**Policy**
0.5	*30*	*0.57*	-	**30**	-	$-\infty$	**2**	-	**60**	**19**	*3*	*19*	*Four-index*
1	*30*	*1.04*	-	**30**	-	$-\infty$	**3**	**21**	**59**	**19**	*4*	*21*	
⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	
3	*30*	*3.01*	-	**30**	-	$-\infty$	**5**	**24**	**57**	**19**	*6*	*24*	*Five-index*
⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	
5	*30*	*5.01*	-	**30**	-	$-\infty$	**9**	**23**	**55**	**18**	*10*	*23*	
5.5	*30*	*5.51*	**23**	**31**	**18**	$-\infty$	*23*	-	-	-	-	-	
⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	*Three-index*
11	*30.09*	*10.87*	**21**	**35**	**19**	$-\infty$	*21*	-	-	-	-	-	
11.5	*30.13*	*11.32*	-	**35**	**20**	$-\infty$	*20*	-	-	-	-	-	
⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	*Two-index*

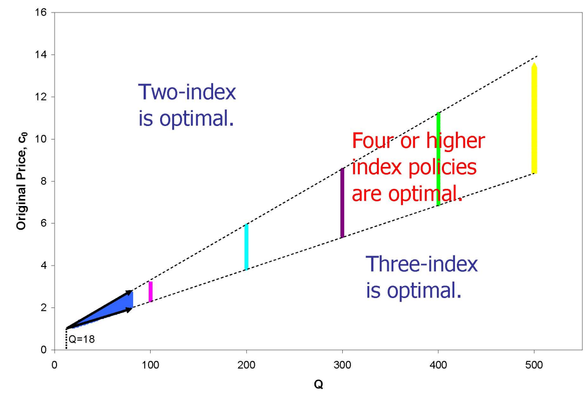
In Example 4, within the buy-and-hold region, the buyer orders only with the discounted price for $\sigma \leq 0.5$. The buyer has nearly perfect visibility into the future and can avoid paying penalty cost without using the original price. However, for $1 \leq \sigma \leq 5$, the buyer ends up having backorders and needs to adjust his inventory by placing small orders with the original price. Outside the buy-and-hold region, the buyer follows a three-index policy for $5.5 \leq \sigma \leq 11$. As we further increase the standard deviation ($\sigma \geq 11.5$), the demand distribution has larger spread and the buyer does not have an incentive to wait additional periods in order to use the discounted price. Therefore, the buyer always orders with the discounted price and two-index policy becomes optimal.

2.4.2 How Bad Can a Three-index Policy Perform?

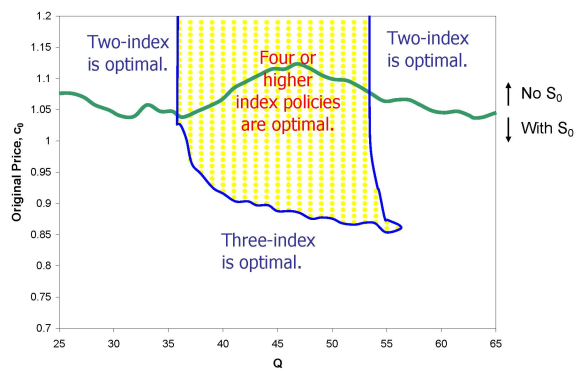
In this section, we study the effect of the basic parameters (Q and c_0) on the buyer's ordering policy. We choose to work on these two parameters as the supplier has more control over these when defining a quantity discount scheme. The non-optimal region may have different shapes and we provide two examples that we have encountered frequently during our analysis.

Figure 2.3(a) displays the structure of the optimal policy for a set of Q and c_0 values. For all values of Q less than 18, a two- or three-index policy turns out to be optimal for all c_0 values. When Q is low but higher than 18, a two- or three-index policy is optimal for most of the c_0 values. However, as Q increases, we observe a larger range of c_0 values where this may not be the case: ordering Q can be considered as procuring for a number of periods and it is only advantageous under a high discount rate. As Q increases, interactions between periods become more important as the decision of receiving the quantity discount has to take the trade-off of holding inventory for longer periods versus procuring cheaper into account. For a fixed c_0, when Q is larger than a threshold the benefit of getting the discount is dominated by the cost of holding inventory. In that case, the buyer uses the discount only when it is necessary; i.e., when his inventory position is very low due to unexpectedly high demands. Observe that the non-optimal region forms a triangle-like connected set.

We repeat the same analysis for a different set of parameter values and display the results in Figure 2.3(b). The shaded region represents the values of parameters where a three-index policy is not optimal. In this example, we observe a line above which S_0 is not observed in the optimal policy, i.e., the buyer does not exercise the original price. Therefore, when a three-index policy is optimal, S_0 is not necessary for policy definition and the actual



(a) Poisson(6), $c_1 = 0.7$, $p = 0.75$, and $h = 0.55$.



(b) Poisson(20), $c_1 = 0.7$, $p = 0.45$, and $h = 0.15$.

Figure 2.3: Structure of the optimal policy for a range of Q and c_0 values.

policy is a two-index policy with parameters (S_1, S_{01}). As the original price no longer has any effect above the line, the policy stays the same and the non-optimal region extends to infinity as we increase c_0 further. Therefore, different from Figure 2.3(a), the region where a three-index policy is not optimal cannot be bounded above with a c_0 and goes to infinity. This observation, which is quite counterintuitive at first, shows that even the minimum order quantity problem may not have a well-defined optimal policy. Hence, a two-index policy is not necessarily optimal as the anomalies due to buy-and-hold continue to occur. There are cases where the buyer orders up to different base-stock levels for different inventory positions. The first row of Example 4 highlights this phenomenon.

In these two examples, a three-index policy turns out to be optimal in a wide range of parameter values. This observation leads us to carry out a more detailed numerical analysis on the performance of a three-index policy. In a numerical analysis, we study 4800 different problem settings with different problem parameters. Table 2.2 summarizes these settings as well as the results of the analysis. For all three distributions, the percentage of the cases where a three-index policy is optimal is steady around $87-91\%$ averaging 88.83%. For the best three-index policy, the maximum deviation from the cost of the optimal policy is 4.3% with a median of 0.025%. As demand variability increases, a three-index policy generally performs much better.

Table 2.2: Performance of the best three-index* policies out of all combinations of $c_0 \in \{1, 1.3, 1.6\}$, $c_1 = 0.7$, $p \in \{0.15, 0.3, 0.45, 0.6, 0.75\}$, $h \in \{0.15, 0.3, 0.45, 0.6, 0.75\}$, and $Q \in \{10, 40, 70, 100\}$.

Discrete Normal† Three-index* optimal: 88.7%			**Poisson** Three-index* optimal: 87.5%			**Discrete Uniform†** Three-index* optimal: 90.7%		
Std. Dev.	**Deviation Average**	**Deviation Max**	**Mean**	**Deviation Average**	**Deviation Max**	**Std. Dev.**	**Deviation Average**	**Deviation Max**
0.5	0.07%	0.25%	10	0.71%	4.31%	2.9	0.44%	1.98%
2.0	0.64%	2.30%	15	0.14%	0.87%	5.8	0.17%	1.25%
4.5	0.31%	1.38%	20	0.18%	0.84%	8.7	0.06%	0.77%
8.0	0.08%	0.86%	25	0.22%	1.12%	11.5	0.06%	0.97%
12.5	0.05%	0.63%	30	0.22%	1.02%	14.4	0.00%	0.06%
18.0	0.00%	0.00%						
Median: 0.038%			Median: 0.040%			Median: 0.012%		

*: Two- or three- index policy †: Demand mean is equal to 30

So far, we gained insights into how a buyer responds to a quantity discount scheme. We have seen that even though a three-index policy is no longer optimal for the infinite horizon problem, it performs exceptionally well under a wide range of problem parameters. Therefore, a three-index policy is an excellent heuristic solution for this rather complicated problem: it is easy to administer and implement, and it captures a high percentage of the profits generated by the optimal policy. Next, we use these results to study the supplier's problem and design efficient quantity discount schemes from her perspective.

2.5 A Closer Look at the Buyer's Orders

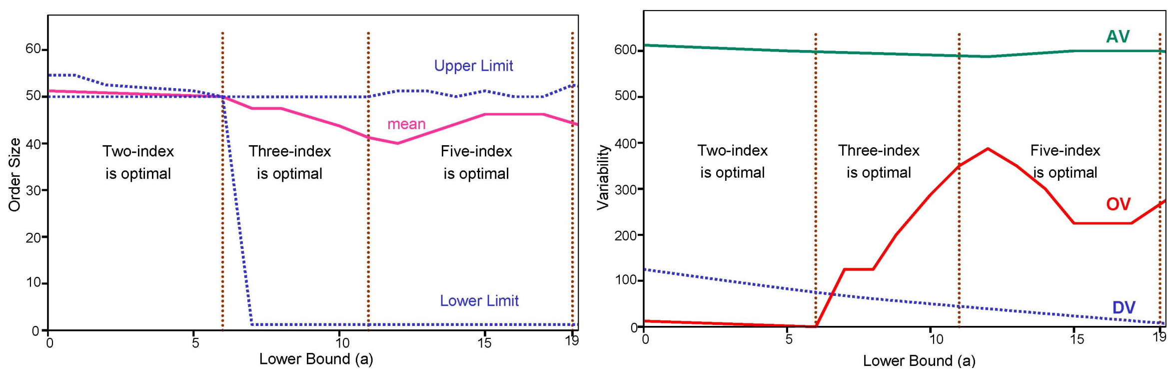
In the previous sections, we discussed the impact of demand variability and the discount scheme on the buyer's ordering policy. Next, we consider the effect of these factors on the orders that the supplier faces. In their 1997 paper Lee, Padmanabhan, and Whang identify quantity discounts as one of the causes of bullwhip effect: the buyer's order batching increases the variability of the orders that the supplier sees. In this section, we question whether decreasing bullwhip effect is always the best alternative for the supply chain in our problem setting[8].

There are different components of variability within a supplier-buyer relationship: demand variance (DV), actual order variance (AV), and order quantity variance (OV). We define AV as the variance of orders placed by the buyer including orders of size 0 and OV as the variance of orders placed by the buyer not including orders of size 0. As buyers batch the orders, AV increases which is known as bullwhip effect. The impact on OV, however, is not that clear.

We know that as the demand variance increases above a threshold the buyer follows policies with fewer indices. What we will do next is to investigate the impact of the policies with fewer indices on AV and OV. Figure 4.4 plots the effect of DV on average order size ρ, AV, and OV. As expected, as DV decreases (i.e., "a" increases), the buyer follows policies with more indices. Furthermore, the supplier's mean order size ρ and actual order size variability AV are much higher than demand mean and variance. By giving discounts, the supplier increases ρ, and decreases the probability of receiving an order in a period.

[8]There are studies with a similar goal in the literature. For example, Chen and Samroengraja (2003) consider a problem where a single product is sold through multiple retailers. In their model, the retailers replenish their inventory from a supplier with finite production capacity. The authors compare two different policies (staggered and (R, T) policies) and show that a replenishment strategy that dampens actual variance does not necessarily reduce the supply chain costs.

Therefore, there are many periods where the buyer does not place an order which inflates AV. Interestingly, AV is quite insensitive to changes in the demand variance. Notice that OV is always less than AV and is *not monotone* as a function of DV.



(a) The effect of DV on order size.

(b) The effect of DV on AV and OV.

Figure 2.4: Demand is Discrete Uniform $(a, 40 - a)$, $c_0 = 1.0$, $c_1 = 0.7$, $p = 0.4$, $h = 0.15$, and $Q = 50$. Bounds are provided for 95% confidence level based on $10,000$ experiments for each a.

In order to further observe the effect of DV on OV, we change the demand variance by keeping its mean constant over time and display the results in Figure 2.5. We observe that the order sizes show a very complex behavior. Small shifts in demand variance cause large shifts in the order variance as the buyer's optimal policy changes abruptly. Similar to Figure 4.4, for larger DV values, the buyer follows a two-index policy and OV is less than DV. However, when DV gets smaller the buyer starts to follow higher index policies where OV dramatically increases and is no longer robust to changes in DV. Therefore, *OV and DV move in opposite directions*. When DV is small the buyer can see the future. Therefore, the buyer becomes more strategic and follows buy-and-hold and wait-and-see strategies with high indices. This phenomenon, which has not been previously studied, is significantly different from the bullwhip effect.

In these examples, the quantity discount seems to make the problem even more complicated by increasing the variability in the system if the buyer does not follow a policy

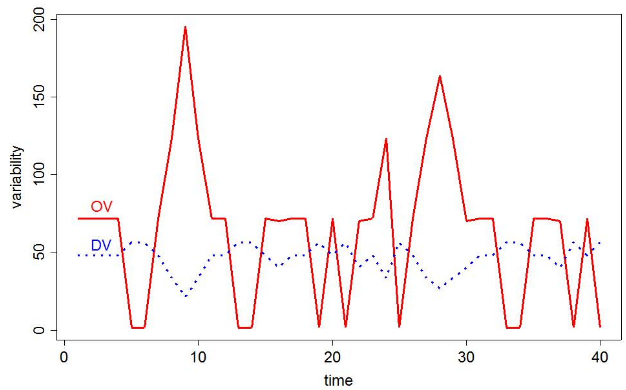


Figure 2.5: The effect of the demand variance on the order variance. Demand is Discrete Uniform with mean of 20, $c_0 = 1.0$, $c_1 = 0.7$, $p = 0.4$, $h = 0.15$, and $Q = 50$. We change the demand variance slightly while keeping the mean demand the same. Notice that OV changes dramatically and in the opposite direction.

with small number of indices. The supplier may actually increase the load on her system by providing an inefficient discount scheme. If the supplier wants to eliminate small orders, a discount scheme with a low OV and high Q can achieve this objective. However, the discount provided can surpass the savings from the discounts. Next, we take a closer look at the supplier's problem and discuss how a supplier can design quantity discounts to maximize her profit.

2.6 Design of Quantity Discounts

As we discuss in the Section 2.1.2, the supplier pays a fixed cost of K for each truckload shipment. In order to avoid paying high transportation cost, the supplier provides quantity discounts to the buyer to increase his order size. Hence the supplier's profit is

$$B(Q, c_0, c_1) = \sum_{n=1}^{N} \left(q_n c_o - K \left\lceil \frac{q_n}{C} \right\rceil - q_n (c_0 - c_1) I_{\{q_n \geq Q\}} \right). \tag{2.14}$$

where $Q \leq C$ and $0 < c_1 \leq c_0$.

The optimal solution for this equation is to set $c_1 = \infty$. However, in practice one of c_0

or c_1 is determined by the market conditions and we do not observe infinite prices. There are two ways to model the problem:

1. When the supplier provides quantity discounts to the buyer to eliminate less than truckload orders, the supplier fixes c_1 and charges $c_0 - c_1$ as an upcharge to cover the extra cost of transportation. In this case c_0 is the decision variable for the supplier.

2. If the transportation costs are not the main driver for the supplier, the market conditions can determine c_0. The supplier provides $c_0 - c_1$ as a discount to increase the buyer's order size. In this case, c_1 is the decision variable for the supplier.

Both cases above create a quantity discount scheme with $0 < c_1 \leq c_0$ and generate a similar behavior from a buyer. The only difference is in the interpretation of the problem from the suppliers perspective. Our results are valid for both cases. In the rest of the chapter we consider the first case in our discussions as it is more common in practice.

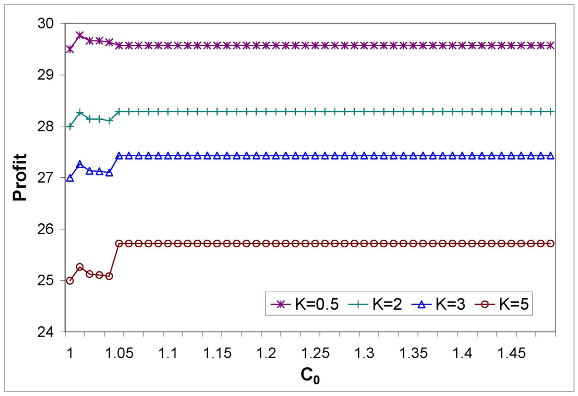
Since there is always uncertainty in the demand and buyers respond to quantity discounts by changing their ordering policies, design of quantity is a challenging problem for the supplier. For example, Heinz considers decreasing $Q = 42,000$ lbs. to motivate the buyers with mean demand around $30,000$ lbs. to increase their order sizes. However, decreasing Q has the risk of providing too much discount which is not covered by the additional savings from the transportation. Therefore, a quantitative approach is necessary for decision support to avoid any unexpected consequences of quantity discounts; hence, we jointly study the supplier's and buyer's problem.

The supplier determines the quantity scheme that would maximize her profits: the revenues are due to the sales of the product and the upcharge that the supplier charges to the buyer and the costs are due to transportation. The analysis of the buyer's problem helps us to understand the response of a buyer to any discount scheme and makes it possible to compute the profit due to a discount schedule. We search over the values of Q and c_0 to find the discount scheme with the maximum profit. As this problem builds on the dynamic programming formulation of § 2.3.2, it is quite difficult to study the supplier's problem analytically. Hence, we depend on numerical analysis (§ 2.6.1) except for some special cases that we study in § 2.6.2.

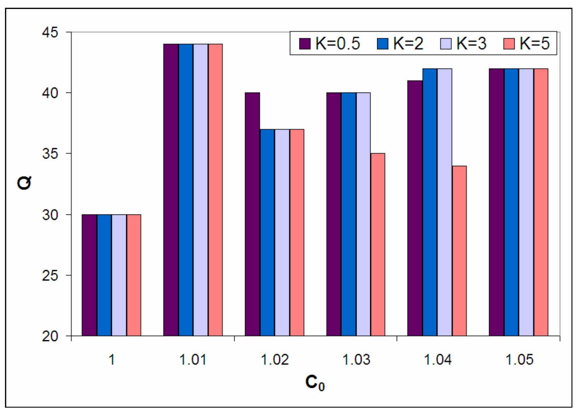
2.6.1 Numerical Analysis of the Supplier's Problem

The increasing transportation costs means higher fixed costs for the suppliers. Fixed costs become a critical factor that affects the decision of the supplier. We study the effect of increasing fixed costs on a quantity discount provided by the supplier in Figure 2.6. For any c_0 value, we provide the optimal profit by finding the best Q for that value of c_0. The values used are in thousand lb units. We use a truck capacity of $C = 45$ and $c_1 = 1$ for each thousand lbs. which are also common in practice. When K is low ($K = 0.5$), the supplier's optimal discount scheme is ($Q = 44, c_0 = 1.01$). At his optimality, the buyer uses three-index policy. As K increases, the supplier's optimal discount scheme ($Q = 42, c_0 = 1.05$) is such that at his optimality the buyer uses a two-index policy. Hence, if it is not economically feasible for the supplier to design discounts with two-index policy (which would decrease the OV as we have seen in § 2.5), the supplier can still use discounts to increase ρ by setting a $Q > 0$ and decrease the number of orders placed by the buyer. By adjusting the discount rate and Q, the supplier can design profitable discount schemes with higher number of indices at the buyer's optimality. When K is low, the system is mostly driven by the inventory related costs of the buyer. Supplier can still tolerate some less than truckload orders with three index policy and provide small discounts to increase the average size of the orders. When K is larger, the system is driven by high fixed cost and the supplier has a higher upcharge to eliminate less than truckload orders.

In this example, for all K values, the supplier is better off by providing some form of quantity discount to the buyer. It is noteworthy that all the discount schemes in this example are found by maximizing the profit functions of the supplier under some different conditions. If the supplier chooses a discount scheme without any quantitative support, the quantity discounts can be far worse off than not providing quantity discount. When there



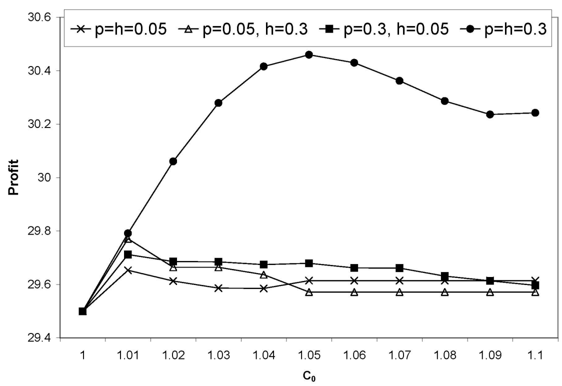
(a) Total profit



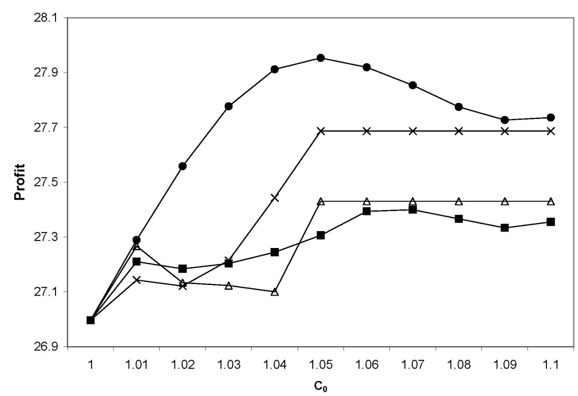
(b) Optimal Q values

Figure 2.6: The supplier's optimal discount scheme for $C = 45$ and different K values. Demand is Discrete Normal$(30, 5)$, $c_1 = 1.0$, $p = 0.05$, and $h = 0.30$.

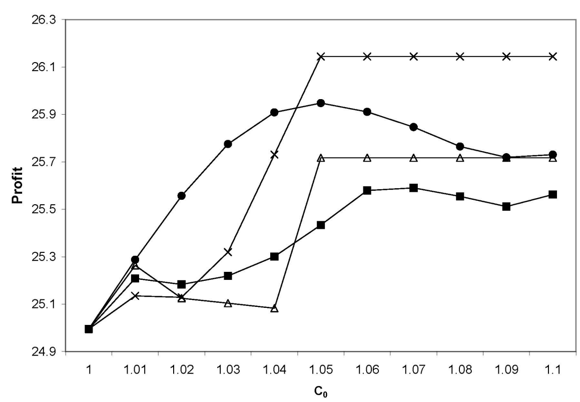
is no discount available, the buyer orders every week and we have $\rho = \mu$. When the supplier sets a $Q > 0$ for the discounted price, the buyer does not place an order in some of the periods. This increases ρ and we have $\rho > \mu$. This also means that order frequency decreases and the supplier incurs lower fixed ordering cost. Even for low K values, incurring lower fixed costs is important for the supplier. Furthermore, the optimal Q is less than the truck capacity, which is 45. Even though the supplier may not use some of the truck capacity to its fullest extend, the difference between the truck capacity and Q creates cushion against the changes in buyer's order size and eliminates the possibility of shipping an additional truck due to large order sizes. A low Q also has a positive side effect by increasing the attractiveness of any quantity discount scheme.



(a) $K = 0.5$



(b) $K = 3$



(c) $K = 5$

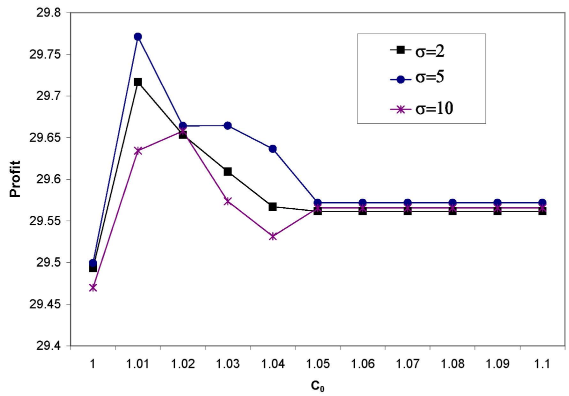
Figure 2.7: The supplier's optimal discount scheme for different p and h values. Demand is Discrete Normal$(30, 5)$ and $c_1 = 1.0$.

Figure 2.7 studies the change in the supplier's optimal discount scheme as the inventory related costs change. When inventory related costs are high and fixed costs are low the system is driven by the inventory costs of the buyer. The buyer places orders which are independent of the quantity discounts due to high inventory related costs. Therefore, the supplier can collect upcharges from the buyer for the orders less than Q. As the supplier increases the upcharge, the buyer does not change his behavior and pays higher upcharge. Therefore, the profit of the supplier increases linearly with the upcharge. However, as the upcharge gets higher than some critical value, the buyer starts to change his ordering behavior. He orders less with the original price to avoid paying high upcharges. Therefore, the supplier increases her profits by collecting more upcharges when the inventory related costs of the buyer is high. When inventory related costs are low compared to the upcharge, the buyer orders with the discounted price. As we increase the fixed costs, we can see that inventory related costs become less important and the supplier provides discount schemes with two-index policies.

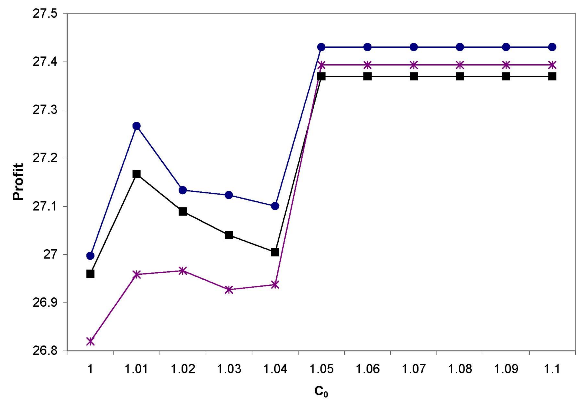
Figure 2.8 studies the change in the supplier's optimal discount scheme as the demand variance changes. As can be seen from Figure 2.8(a), there is no clear relationship between the supplier's profit and the demand variance. The supplier can manage the system better for moderate values of variance; this helps her control the buyer's order more tightly and avoid the problems due to buy-and-hold (when σ is low; Example 4) or extremely erratic demand behavior (when σ is high) that would force her to make inefficient shipments.

2.6.2 Minimum Order Quantity Problem

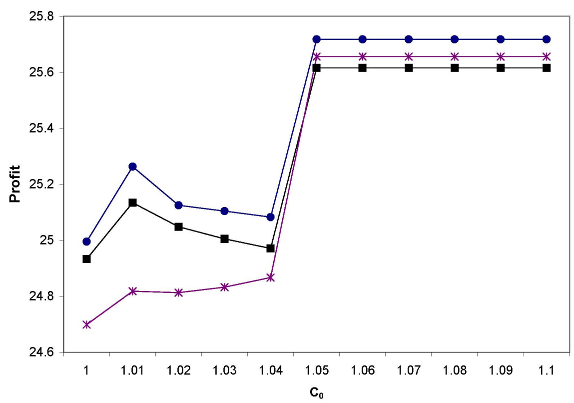
As we discuss in the previous section, increasing fixed costs are forcing the suppliers to totally eliminate less than truckload order. As industries move from low fixed cost to high fixed costs, the suppliers tend to provide quantity discounts with two-index policies. In this case the problem reduces to a minimum order quantity where the buyer always orders at least Q units. However, minimum order quantity cannot by itself guarantee a nice ordering behavior from the buyer. Consider a case where c_0 is very large. In this case, the buyer always orders with the discounted price. From Figure 2.3(b) we know that the buyer can follow a buy-and-hold strategy and even though he never exercises the original price, the



(a) $K = 0.5$



(b) $K = 3$



(c) $K = 5$

Figure 2.8: The supplier's optimal discount scheme for different σ values. Demand is Discrete Normal$(30, \sigma)$, $c_1 = 1.0$, $p = 0.05$, and $p = 0.3$.

optimal policy is not necessarily a two-index policy. However, when the supplier sets Q large enough, she can eliminate this possibility:

Theorem 3 *When c_0 is very large, there exists a finite Q^* such that:*

i. For $Q > Q^*$ the buyer follows a two-index policy with indices $(S_1,\ S_{01})$. The optimal *order quantity* $q^*(x)$*:*

$$q^*(x) = \begin{cases} \max\{S_1 - Q, Q\} & \text{when } x \le S_{01} \\ 0 & \text{when } x > S_{01} \end{cases}$$

ii. As Q *increases,* S_{01} *converges to* $\hat{S}_{01} = \mu - Q\left(\frac{h}{h+p}\right)$ *and the buyer's average cost converges to* $C(Q) = \frac{Qhp}{2(h+p)}$*.*

Proof: The proof is available in an on-line addendum, Section 5.1.4. □

The supplier can choose any $Q > Q^*$ in order to make the buyer follow a two-index policy. When the buyer does not use a buy-and-hold strategy for any Q, Q^* is simply equal to 0: no matter how low the value of Q^* is, the buyer will always follow a two-index policy (Figure 3(a)). However, the buyer should also look for the optimal level of discount in order to eliminate the small orders.

Theorem 4 *In a quantity discount problem, for fixed and finite values of* Q*,* p*,* h*, and* c_1*, there exists a discount* $z^*[= 1 - (c_1/c_0)] < 100\%$ *such that,*

a. for every discount ratio $z \ge z^*$*, the buyer does not exercise price* c_0 *for any initial inventory position;*

b. for every discount ratio $z < z^*$*, the buyer exercises both prices* c_0 *and* c_1*.*

Proof: Available in an on-line addendum, Section 5.1.5. □

In Theorem 3 we show that by adjusting Q the supplier can guarantee that the buyer follows a two-index policy for larger values of c_0. Actually, the supplier can control the buyer's ordering behavior even more tightly. Let α be the percentage of the orders which are strictly greater than Q when the buyer follows a two-index policy. In Figure 2.9, for different levels of Q, α values are displayed. For these Q values which are greater than $Q^* = 55$, the buyer does not buy-and-hold and the optimal policy is a *two-index policy* (Theorem 3). As Q increases, α decreases dramatically. For even a moderate value of Q

such as $Q = 60$ (3 weeks' demand), the probability of an order greater than Q is less than one in a thousand. The optimal policy when $Q = 60$ is a two-index policy with indices $(S_1 = 36, S_{01} = 4)$. For all practical purposes, the inventory level almost never falls below $S_1 - Q = -24$ and the buyer almost always orders exactly Q. In this case, the critical level where the buyer stops ordering S_{01} is very close to $\hat{S}_{01} = 5$ and the critical level becomes independent of demand variability.

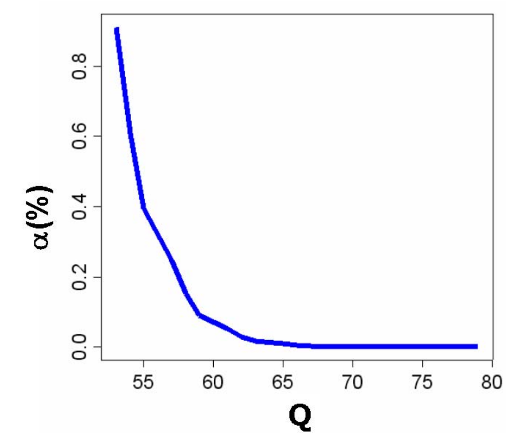


Figure 2.9: Values of α for $Q > Q^* = 55$. Demand is Poisson(20), $p = 0.45$, and $h = 0.15$.

As is the case without α, we can find discount ratios that would make a buyer follow a two-index policy and place orders of size greater than Q only α percent of the time (an (α) two-index policy). Figure 2.10 displays $(Q, z^*(Q))$ when $\alpha = 0.001$. Note that the optimal discount to be provided is not always monotonic in Q. It is possible for the supplier to decrease the discount and increase Q at the same time, which seems to be counterintuitive at first. Remember the two policies that we discussed in Section 2.4.1. When Q values decrease, the buyer does not need to follow a *wait-and-see* strategy as even when he procures with the original price in a given period, the chances that he will get the discount in the upcoming period will be high. Therefore, the buyer will start to follow higher index policies by utilizing the original price. On the other hand, when the discount rate is high, a *buy-and-hold* strategy becomes even more attractive. This time the buyer does not use the original price, but he starts to use higher index policies with the discounted price in order to utilize the discount opportunity. In between, we may see dynamics that balance these two extremes and lead to a two-index policy even when the discount opportunity is inferior.

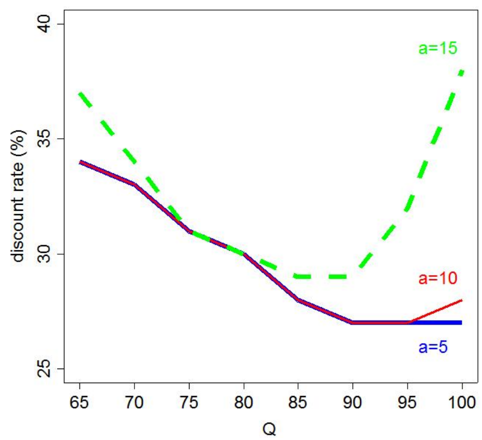


Figure 2.10: Minimum discount percentages $z^*(Q)$ that guarantees an (α) two-index policy. Demand is Uniform $(a, 40-a)$, $p = 0.45$, $h = 0.15$, and $\alpha < 0.001$.

Another interesting effect that we observe in Figure 2.10 is the impact of demand variance. *The higher the demand variance the less discount the supplier needs to provide.* In less variable systems the buyer is more likely to buy-and-hold (Example 4). Therefore, in these systems, in order to regulate the buyer's ordering behavior, the supplier should provide more attractive quantity discounts; i.e., given a Q, she should give bigger discounts.

Hence, when the setup cost is very high, the operational objectives of the supplier may be in parallel with observing a small OV from the buyer with a high ρ. This way, the supplier would guarantee (with high probability) full or near full truckloads. For the our objective function, order variance is more crucial for the supplier compared to the actual variance.

2.7 Conclusion and Future Research

There are many potential advantages of quantity discounts in a supply chain. However, unless enough care is given to the design of a discount scheme much of the potential benefits could be lost. In this paper we present a model that captures the buyer behavior under stochastic demand when the supplier provides quantity discounts. Based on the optimal response of the buyer that we have analyzed in detail, we suggest strategies for the supplier in designing discount schemes.

Our paper shows that **many times operational difficulties in a quantity discount setting are self-created**. Policies with a high number of indices can be optimal in the presence of quantity discounts in a stochastic environment largely due to the supplier's choice of parameters. This may lead to orders that can be very difficult to manage for the supplier. By choosing discount schemes carefully, however, the supplier can moderate the orders effectively. Contrary to the well-studied bullwhip effect problem, in this setting **the erraticity of the orders may increase while the basic demand variability decreases**. While a three-index policy is optimal for the buyer in many cases and near optimal in all cases, it can still create high order variance, especially when it is not carefully designed. We define a cost function for the suppliers and show that higher fixed costs are forcing suppliers to eliminate less than truckload orders. A well designed two-index policy, if profitably enforceable, is to the supplier's greatest advantage.

Additional issues such as time windows and volume discounts should also be analyzed. In our model, the supplier stabilizes the orders of the buyer by designing an effective quantity discount scheme. However, this has the effect of increasing the time between orders which reduces the order variance of the supplier. Cachon (1999) discusses that suppliers should not take actions that can increase the buyer's cost substantially in order to dampen their order variabilities. In our analysis, the supplier does not put any restriction on the buyer about the timing of the orders. The buyer voluntarily orders with the discounted price and is better off with respect to the system without a discount. The supplier can adjust the discount rate and Q in order to control the order size and variance of the buyer. There is an increase in the holding and penalty costs of the buyer, which is compensated by the discount rate. Therefore, the supplier shares the benefits of reducing order variability with the buyer.

In addition to increasing the time between orders, the efficient quantity discount scheme that we propose shifts the variability of the system from order sizes to order timing. A possible extension of our analysis would be to study how to best moderate the variation in number of periods between orders as well. Another extension that we work on includes multiple buyers with significant heterogeneity. In this case, the supplier may provide different price breaks such that each buyer picks exactly one of the breaks at his optimality. Since each buyer only exercises one price break, the supplier happens to price discriminate her customers based on their end customer demand distribution. Finally, we would like to emphasize that there are many ways to model the quantity discount problem. Our model captures the most common practices in industry.

Chapter 3

A Data Mining Approach to Forecast Behavior of Manufacturers in Automotive Industry

3.1 Introduction

The rapid expansion of computer resources creates the potential to bring business intelligence into decision support systems of supply chains. With the increased availability of enterprise wide databases, the amount of data collected is growing at a tremendous rate. One objective of supply chain management is the reduction or elimination of all activities that do not add value and concentrating on factors that maximize value and productivity. Collaborative Planning, Forecasting, and Replenishment (CPFR) is an evolution and refinement of these concepts among the players in the supply chain. CPFR is an initiative intended to improve the relationship among all participants in the supply chain through jointly managed planning and shared information. The quality and the intensity of the information exchange demands a strong commitment of cooperation from the participating organizations. The planning and forecasting components require intensive information exchange in all levels of relationships. Unsatisfactory relationships between the parties lead to inefficient information flow in the supply chain.

We consider a customer order prediction model in which forecasts of orders at some future date are used as input to a series of inventory planning decisions. We are analyzing the orders that are received by a supplier who produces parts for auto manufacturers. Multiple customers (auto manufacturers) requesting multiple parts, place preliminary orders which are estimates for actual orders starting from six months before the due date. Customers can update their orders as the due date approaches. The supplier guarantees very high service levels as stipulated in the contract with the customers. Since capacity is limited and there is lead time for production, the supplier may fail to fulfill orders. Therefore, production planning and capacity decisions have to be made carefully in order to achieve high service levels. As the production quantities are being committed so far in advance, it is very difficult to predict final quantities of each part desired. Our goal is to improve the supplier's operations through a better understanding of her customers' ordering behavior.

We provide a methodology that can be used in order to analyze the forecast behavior of manufacturers. In the first step of the analysis, manufacturers' forecast data are transformed into a format that can be used as input for further analysis. Orders are replaced by daily flows in order to predict the daily requirements of the manufacturers and overcome any complexities due to ordering problems such as order-splitting, order-combining and changes in due date. In the second step of the analysis, we use data mining techniques such as clustering and projection methods in order to visualize forecast behavior of the manufacturers.

In our analysis, we introduce different order complexities that have not been discussed in literature. Analytical models that assume similar ordering behaviors can be used to obtain policies that improve the supplier's operations. Another important point is the lack of empirical research in the area of supply chain management. Our results provide strong empirical support for models that assume different forecast behaviors. In our analysis, we show that customers are consistent with their forecast behavior over time. It is possible to design different incentive mechanisms that improve a customer's performance. One interesting result is under-estimation being a more common behavior than over-estimation. There are models in literature that explain inflation of order sizes under different assumptions (Cachon and Lariviere 1999a, 1999b). However, the reasons behind under-estimation is not well studied in literature. From the automotive industry dynamics reputation is the most likely cause for customers to under-estimate their orders. Therefore, the models that consider longterm relationships should take reputation into account.

3.2 Environment Analysis

3.2.1 Automotive Suppliers

As auto manufacturers concentrate more with the customer side, production and engineering move more to suppliers. Auto suppliers develop and build 65% of the average vehicle today; according to a recent study this number is expected to be 77% over the next decade as suppliers provide more engineering and production of the components. Suppliers are the main part of the value creation in the auto industry. In 2002, top 100 global automotive suppliers delivered $366.4 billion in parts to auto plants around the world, up from $333.8 billion in 2001. Companies such as Bosch, Continental, Delphi, Lear, Siemens VDO Automotive, ThyssenKrupp, and Visteon are expected to increase their value creation 70% by 2015. Although globalization is a well established trend among the industry's biggest players, many of the largest suppliers still remain heavily rooted in their home markets. High customization of products and geographically dispersed operations require an efficient supply chain management for auto parts. Decreasing margins in industry and high technology requirements result in very demanding auto manufacturers. For example, supplying Honda requires patience. In some cases, the auto manufacturer would talk with a potential supplier as long as two years before deciding to offer a contract. Larry Jutte, head of purchasing for Honda of America Manufacturing, Inc. comments about their search for suppliers in the North American market: "We began by trying to identify suppliers in North America that matched our core values. Some thought Honda's value system was more challenging, they thought it was unreasonable. There is no question that we were demanding about quality, cost and delivery, we wanted to know a heckuva lot more about a supplier than most other auto companies at that time." Therefore, a strong tie and collaboration is a key to lengthy and profitable relationships in the automotive industry.

3.2.2 Order Forecasts

Order forecasts are an essential part of collaboration when the capacity is limited and there is lead time for production. Forecast is the basis for the integration of a manufacturer to the production process. In CPFR, even sharing point of sales data is not enough for effective and efficient production planning. Sharing demand information alone will not guarantee uncertainty reduction in the system. One barrier is the time horizon. Since there is lead time and orders are placed in fixed epochs due to fixed costs, manufacturers should provide early forecasts.

There are two components of variation in the system. First one is the noise which comes from the uncertainty of the nature. The current demand becomes a better predictor for the forecasted demand with the proximity to actual delivery. A forecast for tomorrow will be more accurate than one for next month. The latter is the manufacturer's bias in terms of over-estimating or under-estimating the order size. Manufacturers can improve their forecasting performance and can reduce the noise. However, as long as it is not incentive compatible to submit true forecasts, manufacturers will transform their forecast and truthful transmission of forecast is not possible. In our analysis, we provide empirical support for different forecast behaviors for the manufacturers.

3.2.3 Database Information

In our analysis, we are analyzing the forecasts and orders that are placed by multiple customers (auto manufacturers) to a supplier. The supplier is one of the biggest engine systems suppliers in the automotive industry. We provide an analysis for his “hot” selling and high investment ($5.3 billion) new engine system. Multiple customers requesting multiple engine parts place preliminary orders that are estimates for actual orders starting from six months before the due date. Customers can update their orders as the due date approaches.

Our database consists of the orders placed by 497 customers for 35,551 different parts in years 2000 and 2001. The total number of orders is more than a million. Each part belongs to one of 214 product families. Parts in product families have similar functions and prices. An average price for a product family is also available in the data.

The attributes for different databases are as follows:

1. **Order database:** customer, part, order date, due date, quantity, warehouse, factory information
2. **Part database:** product family
3. **Product family database:** average price for family

Pareto Result: Customers in our analysis are not identical in total dollar value of their orders. A widespread empirical result about income dispersion, known as Pareto result, implies that 20% of the customers will account for about 80% of the total demand. Ordering the customers in descending revenue in Figure 3.1, 11% (55 customers) of the customers represents 80% of the revenue.

Some of the customers represent $3-5\%$ of total revenue. Each of the top 31 customers' share in total revenue is higher than 1%. Therefore, by analyzing those 31 customers (just 0.6% of customers), we can explain 65% of the total revenue. For the next steps

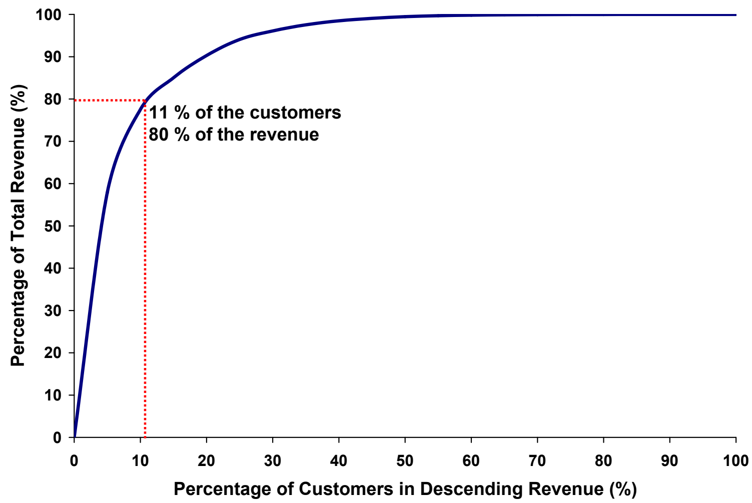


Figure 3.1: 11% of the customers accounts for 80% of the revenue

of our analysis, we separately analyze the major customers and provide statistically more significant results.

3.2.4 Objective

As part of CPFR, the supplier asks customers (auto manufacturers) to submit forecasts in order to start production in advance. However, a customer's forecast accuracy can be very low and forecasts can be very misleading. Our goal is to define a framework for analyzing the forecast performance of the customers and to provide empirical support for different forecast behaviors. In our analysis, we explore and analyze large quantities of data to discover meaningful patterns and rules for the ordering process of customers. We try to explore the following questions:

a. Are classical forecasting techniques applicable? If not, is it possible to clean the forecasts in order to prepare the data for a deeper estimation analysis?

b. Is it possible to provide a quantitative representation of a customer's forecast performance?

c. Can we visualize the behavior of customers through time?

d. Can customers be assigned to groups? Do these groups represent significant commonality among different forecast behaviors?

e. How can we recognize different forecasting patterns? Is it possible to automate the process of anomaly detection from the data?

f. Are customers consistent with their forecast behavior? What is the general tendency in terms of over-forecasting or under-forecasting?

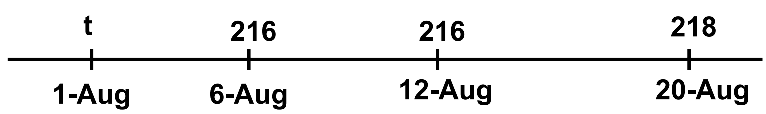
3.3 Literature Review

The management science literature has paid a great deal of attention to forecasting, so we only provide a description of major classes of models related to our analysis.

Most of the literature is based on providing mathematical models for evolution of forecasts. Graves et al. (1986a, 1986b, 1998) and Heath and Jackson (1994) develop the Martingale Model of Forecast Evolution (MMFE) to model how demand forecasts of customers evolve in time. In MMFE, a forecaster generates a demand forecast for a single item in different periods. The forecast update errors, the difference between any two periods' forecast for a future date, are assumed to follow the Martingale Property: forecast update errors are independent, identically distributed, multivariate normal random variables with mean 0. Gullu (1996) , Graves et al. (1998), and Toktay and Wein (2001) use MMFE to model the evolution of the forecasts. Another approach is using Bayesian updates to incorporate new information as it becomes available . Scarf (1959) , Azoury (1985) and Lariviere and Porteus (1999) provide insights into the order evolution problem from Bayesian perspective. Chen et al. (2000) , Aviv (2001) and Aviv (2002) have suggested several other stylized theoretical models to study inventory planning when there are forecasts available for future demand. From the analysis of the customer forecast data, we capture complexities with order updates such as order-splitting, order-combining and shifts in due dates. The demand forecasts do not follow a regular pattern and the data without any transformation does not fit in the classical forecasting models.

The empirical research about forecast behavior of customers is quite limited in literature. Terwiesch et al. (2003) considers the problem from a buyer's perspective. A buyer is placing orders from a set of equipment suppliers. The forecast performance of the buyer is measured according to the forecast volatility (continually updating the orders) and forecast inflations. They demonstrate that inflating forecasts and providing volatile orders damage the buyer's reputation and thereby lead to a lower level of service in the future. In our analysis, we looked at the problem from the supplier's perspective and provide analysis for understanding the forecast behavior of the customers.

3.4 Understanding Customer Forecasts



Example 5 *A customer forecast at order date 1-Aug.*

At each order date t, customer i updates all his previous forecasts and can place some new forecasts for part j. Since we repeat the same analysis for all customers and parts, we drop the indices i and j for the rest of the analysis. Therefore, at order date t, the customer places forecasts for different due dates $S_t = (s_{1,t}, s_{2,t}, ...)$ where $s_{r,t}$ is the r^{th} due date of the order date t. Time between order dates and time between due dates should not necessarily follow a regular pattern. Each customer can provide a forecast update at any date. At t, the customer provides a set of forecasts $F_t = (f_{s_{1,t}}, f_{s_{2,t}}, f_{s_{2,t}}, ...)$ for all days in S_t. The forecast vector F_t is the most recent forecast information at time t and $f_{s_{1,t}}$ denotes $(s_{1,t} - t)$ days advance forecast. The most recent order date provides the valid forecast information for estimation. If there are no more updates for a due date, the forecast is treated as the final order. In Example (5), for t=1-Aug, the due date vector is S_t=(6-Aug, 12-Aug, 20-Aug) and the forecast vector for these particular due dates is $F_t = (216, 216, 218)$.

3.4.1 Order Definition

In our analysis, customers can modify both their due dates and forecast quantities at order dates. Changes in due dates make it difficult to track the update of an order at different order dates. From the data set, we provide major complexities that make classical forecasting tools inapplicable.

Example 6 *12-Aug and 20-Aug orders in 1-Aug order splits into 12-Aug, 16-Aug, 20-Aug and 24-Aug orders in 8-Aug order.*

1-Aug		**8-Aug**	
Due date	**Forecast**	**Due date**	**Quantity**
6-Aug	*216*		
12-Aug	*216*	*12-Aug*	*107*
		16-Aug	*109*
20-Aug	*218*	*20-Aug*	*112*
		24-Aug	*106*

Order-splitting

Some customers substitute their forecasts for a given due date with smaller batches as the due date gets closer. Customers tend to place large preliminary aggregate forecasts and then split them into smaller orders. We observe that a customer can place monthly forecasts six months in advance and change it into bi-weekly forecasts three months in advance and finally end up with weekly forecasts in the last month before the shipment. In Example 6, forecasts for due date 12-Aug and 20-Aug are replaced by 12-Aug, 16-Aug, 20-Aug and

24-Aug forecasts in 8-Aug order. When a forecast gets split, we lose track of this particular order. We have to make extra assumptions in order to obtain the actual order quantity for the order that gets split.

Order-Combining

Some customers combine their orders as the due date gets closer. The initial order no longer exists after combining. Therefore, there is no actual order quantity for that forecast. The accuracy of the order before combining is hard to derive without making any extra assumption about a customer's combining policy. In Table 7, forecasts for 12-Aug, 16-Aug, 20-Aug and 24-Aug orders from 1-Aug order are replaced by forecasts for 12-Aug and 20-Aug orders at 8-Aug.

Example 7 *12-Aug, 16-Aug, 20-Aug and 24-Aug orders in 1-Aug order are combined into 12-Aug and 20-Aug orders in 8-Aug order.*

1-Aug		**8-Aug**	
Due date	**Forecast**	**Due date**	**Quantity**
6-Aug	*110*		
8-Aug	*124*		
12-Aug	*124*	*12-Aug*	*256*
16-Aug	*132*		
20-Aug	*112*	*20-Aug*	*252*
24-Aug	*140*		

Example 8 *8-Aug, 12-Aug, 16-Aug, 20-Aug and 24-Aug orders in 1-Aug are shifted by two days to 10-Aug, 14-Aug, 18-Aug, 22-Aug and 26-Aug in 8-Aug order..*

1-Aug		**8-Aug**	
Due date	**Forecast**	**Due date**	**Quantity**
6-Aug	*110*		
8-Aug	*124*	*10-Aug*	*120*
12-Aug	*124*	*14-Aug*	*128*
16-Aug	*132*	*18-Aug*	*134*
20-Aug	*112*	*22-Aug*	*120*
24-Aug	*140*	*26-Aug*	*134*

Shifts in Due Dates

Customers can modify due dates of orders while updating their forecasts. We observe that it is difficult to keep track of an order when there is a change in its due date. Initial forecast starts six months before the shipment. Therefore, shifts in due dates are inevitable in a manufacturing environment with lots of uncertainty in production and demand sides. In Example (8), forecasts for 8-Aug, 12-Aug, 16-Aug, 20-Aug and 24-Aug orders are replaced by forecasts for 10-Aug, 14-Aug, 18-Aug, 22-Aug, and 26-Aug orders. In Example (8), there is a two day shift in due dates and the customer is also adjusting his forecast quantity.

A New Order Evolution Model

In the MMFE framework, the order vector is updated at each order date. In our data set, the customer does not have to provide an order update at each order date t. The time between order dates with updates does not necessarily follow a pattern. In MMFE, the customer is assumed to provide forecasts at each order date for each period in the planning horizon. Customers can only update the quantities and cannot change the due date of an order in MMFE. However, in our analysis the customers first choose the due dates S_t at order date t and then provide the forecast vector F_t for these days in S_t. The customer does not provide forecasts for all days in the planning horizon. He submits forecasts only for the due dates in S_t. The number of days between the due dates in S_t does not necessarily stay constant. In our analysis, we also observe that customers can shift due dates. Moreover, order-splitting creates new due dates and order-combining reduces the number of due dates. Order-combining, order-splitting and shifts in due dates can happen simultaneously, which makes it difficult to track orders.

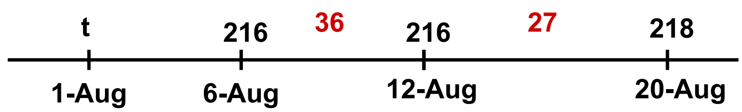
When an order gets split, combined or shifted, having order numbers does not solve the problem of finding an accuracy level for this order. In order to handle these complexities, we look at the periods in which the customer consumes these quantities. We assume that customers consider their daily production requirements while placing orders. A systematic approach is disaggregating the order quantities and work at daily flow level.

3.4.2 Daily Flow Analysis

At each order day customers aggregate daily requirements and place orders for some due dates which are discrete points in the planning horizon. In our daily flow analysis, we try to predict the daily requirements of a customer based on the forecasts. There are different ways of estimating daily flows from orders. The greedy solution is assigning any order to the following days before the next due date. So at order date, we divide the orders $f_{s_{k,t}} \in F_t$ by $s_{k+1,t} - s_{k,t}$ and assign it evenly to the days between $s_{k+1,t}$ and $s_{k,t}$. We assume that $f_{s_{k,t}}$ is consumed during those days. Other smoothing techniques can be considered with extra assumptions. In our analysis, we assign daily flows after a due date unless the order is the latest due date of an order forecast vector F_t. Otherwise, we cannot find the days when the quantities ordered at the last days of an order date get consumed. Example (9) demonstrates the assignment of daily flows for order date 1-Aug.

Handling order-splitting, order-combining and shifts in due dates is important for performance analysis of the customer. However, our operational objective is to understand how customers change their order quantities, including the orders which do not have any complexities. We provide a procedure to find a quantitative representation for each customer's forecast performance:

Example 9 *At order date t=1-Aug, 216 is ordered for 6-Aug and it will get consumed between 6-Aug and 12-Aug (6 days), so the daily flow between 6-Aug and 11-Aug is* $216/6 = 36$*. Between 12-Aug and 19-Aug, the daily flow is* $216/8 = 27$ *following the same argument.*



Order date	Due date	Forecast	Number of days	Daily flows
1-Aug	*6-Aug*	*216*	*6*	*36*
	12-Aug	*216*	*8*	*27*
	20-Aug	*218*		

1. **Finding the Accuracy Ratio of Forecasts**

   By running the daily flow analysis on the forecast update at order date t, we can obtain the daily forecasts ($\hat{f}_t^m$) for each day m in the planning horizon at order date t. By repeating the same daily flow analysis on the firm orders, we can generate the actual daily flows d_m for any day m. By having the actual daily flows and the daily forecasts, we can determine the accuracy ratio of an r-day advance order. Different time windows can be considered for accuracy calculations. In our analysis, we consider 30-day advance orders since most customers provide forecasts at least 30 days before the due date. The same analysis can be repeated for other r values. One major complexity in finding the accuracy level of a 30-day advance order is defining a 30-day advance order. It is not often the case that there exists a particular order which is placed exactly 30 days from the order date. One solution is comparing forecast daily flows and actual daily flows for the 30-day in advance. However, this is very sensitive to small shifts in due dates. A more robust solution is considering the total orders in a time window. From the data, we observe that customers provide orders on a weekly basis as due date gets closer. Therefore, we combine daily flows between 30 and 36 days (a week of orders) as a single order and compare it with the sum of actual daily flows for these 7 days. So the accuracy ratio ϕ_r^t of r-day advance order at order date t is

$$\phi_r^t = \frac{\sum_{n=r}^{r+6} \hat{f}_t^n}{\sum_{n=r}^{r+6} d_n}. \tag{3.1}$$

2. **Taking the Log of the Accuracy Ratios**

   It has been discussed in literature that using log transforms for ratio-scaled data, has more accurate results. (Armstrong (2000)) and Hausman (1969) transformed the order evolution problem into a finite horizon sequential decision problem using a quasi-Markovian or Markovian model. They provide empirical support (not entirely) for ratios of successive forecasts being independent and having a Lognormal density function conditional on the change in the forecast. Therefore, we take the log of the accuracy ratios in our analysis in order to give the same weight to under- and over-estimation of the orders.

3. **Assigning Orders to Bins**

   In order to represent the behavior of a customer, we create a histogram for each customer by looking at 30-day advance orders. We assign orders to bins (different intervals in the histogram) by looking at the log value of the forecast's accuracy ratio. We consider 15 different bins with equal widths. There are three groups of bins: one true-estimator (accuracy ratio ranged from 80% to 125%), seven under-estimators (having ranges scaled being multiples of $\log 0.8$) and seven over-estimators (having ranges scaled being multiples of $\log 0.8$).

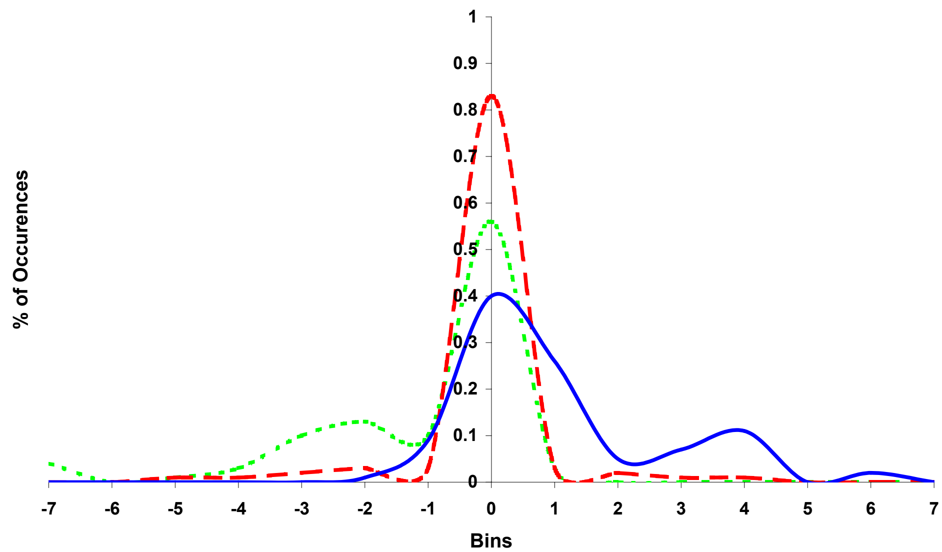


Figure 3.2: The dashed curve corresponds to a customer who falls into the true-estimator bin 85% of the time. The solid curve exhibits the behavior of a customer who over-estimates the size of his orders most of the time. The dotted curve represents the ordering pattern of a customer who under-estimates the size of his actual orders.

For each customer, we start to fill in the bins as the daily flows are calculated. Based on the log value of the accuracy ratio the appropriate customer bin is increased accordingly. The customers can replenish different parts with different quantities. Therefore it is reasonable to have an amount of increase which is equal to the dollar value of an order. After completing assignment of orders to bins for a customer, we normalize the histogram and find the relative frequency of each bin. So the value of each bin corresponds to the ratio of total dollar value of 30-day advance orders that fall into this accuracy level. We can use this 15-dimensional vector to mimic the distribution of a particular customer and input it as a performance measure in data mining. In Figure (3.2), resulting order distributions for 3 different customers are shown.

As a result of our analysis, we have a *customer order distribution* p_i^q for customer i and quarter q. We consider quarters of 2000 and 2001 in our analysis. Each p_i^q is a vector of size

15. The first 7 components of p_i^q(indexed from -7 to -1) corresponds to under-estimation values. In Figure (3.2), 0 on x axis which is the 8^{th} component of the vector p_i^q represents the probability of true-estimation and the last 7 components of p_i^q (indexed from 1 to 7) shows the over-estimation probabilities.

3.5 Characterizing Customer's Forecast

An important issue for a supplier with multiple customers is to provide performance benchmarks. Customer forecast data contain lots of variables which make it difficult to make comparisons among the customers. In Section 3.4, we provide the methodology to derive a quantitative representation of a customer's forecast performance. In this section, we provide a distance metric in order to compare customer order distributions of customers. We then provide supervised and unsupervised clustering techniques that form performance groupings among customers.

Clustering heuristics assign each observation or object to a group. We can cluster the customers by looking at their order distributions. If a customer clusters with over-estimators or under-estimators, then he must be treated with care on the grounds that there is a problem with his order process. If he clusters with true-estimators that concern disappears. We can observe the evolution of a customer's forecast behavior by looking at his cluster for various quarters. We can call the cluster, which includes the customer with 100% forecast accuracy as the ideal cluster, and explore the other clusters based on their distance from the ideal cluster.

In data mining, the main goal is to produce simplified descriptions and summaries of large data sets. As long as there are only two or three dimensions it is easy to visualize two- or three-dimensional graphs. However, as the dimensionality of the data gets larger, it gets difficult to plot a vector of relationships between different factors. Projecting high-dimensional data sets as points on a low display (usually two-dimensional) is a one way of visualizing the data. In our analysis, we project the clusters on a two dimensional space.

A simple clustering is dividing customers into three groups: over-estimators, true-estimators and under-estimators. However, there are advanced clustering techniques which provide better insight. We apply two clustering techniques (K-Medoid Analysis and Self-Organizing Maps) in our analysis. For projection two two-dimensional display, we consider Sammon's Mapping which finds a mapping where the distances between the image points of the data items remain similar to distances in the original metric. Sammon(1969) describes a nonlinear mapping algorithm which has been found to be highly effective in the analysis of multivariate data. The special feature in Sammon's Mapping is that errors are normalized by distances in original space. In our analysis, we consider data for different quarters of 2000 and 2001 for each customer. By looking at the trajectories of the customers on the mapping, customers can be informed about their performance in order to improve their ordering process and offer incentives.

3.5.1 Comparing Customer Performances

Customers can be represented in terms of proximity between their performance vector. In order to compare customer order distributions, we define a distance measure that provides a good approximation of the proximity between customers.

The distance measure should be compatible with the problem characteristics. The vector of customer distribution(p_i^q) has some distinct properties. Each component corresponds to a discrete probability and the sum of the mass probabilities adds up to one. The order of the vector is important and should be considered in the selection of the distance measure. The following distance measure is used in the rest of our analysis.

Definition 4 *Having the customer probability vector p_i, a distance $d(i, j)$ between customers i and j can be defined as follows,*

$$d(i,j) = \sum_{k=-7}^{k=7} \parallel \sum_{t=-7}^{t=k} p_i^q(t) - \sum_{t=-7}^{t=k} p_j^q(t) \parallel \qquad (3.2)$$

where $p_i^q(t)$ *denotes the* t^{th} *component of vector* p_i^q *and* $\parallel . \parallel$ *stands for euclidean metric.*

Our distance measure provides the distance between the cumulative probabilities. Since each distribution is an ordered vector, the distance measure should depend on the order of the vector. Our distance measure is robust to small errors in the calculation of the mass probabilities and puts more weight on the cumulative probability of any point. Also, the distance is not skewed to over-estimation or under-estimation. The distance between two order vectors stays the same if we transpose the order vectors. We use our distance measure to form customer clusters which have similar forecast behavior.

3.5.2 Customer Clustering Analysis

The goal of clustering analysis is to partition the observations into groups so that pairwise dissimilarities between those assigned the same cluster tend to be smaller than those assigned in different clusters. Each observation is assigned to one and only one cluster. The objective of our clustering analysis is to describe forecast behaviors. Each cluster represents customers who have similar forecast behavior. By looking at the movement of customers between the clusters for different quarters we can track the ordering pattern of a particular customer over time. If a customer shows up in the same cluster for all quarters, it means the customer is consistent with his ordering behavior. Our clustering analysis also automates the process of setting performance benchmarks. Clustering also provides a dynamic measure of the ideal behavior for customers.

K-Medoid Analysis

We introduce K-Medoid clustering technique which is a modified version of the well known K-Means clustering technique. The only difference is having actual customers as the cluster

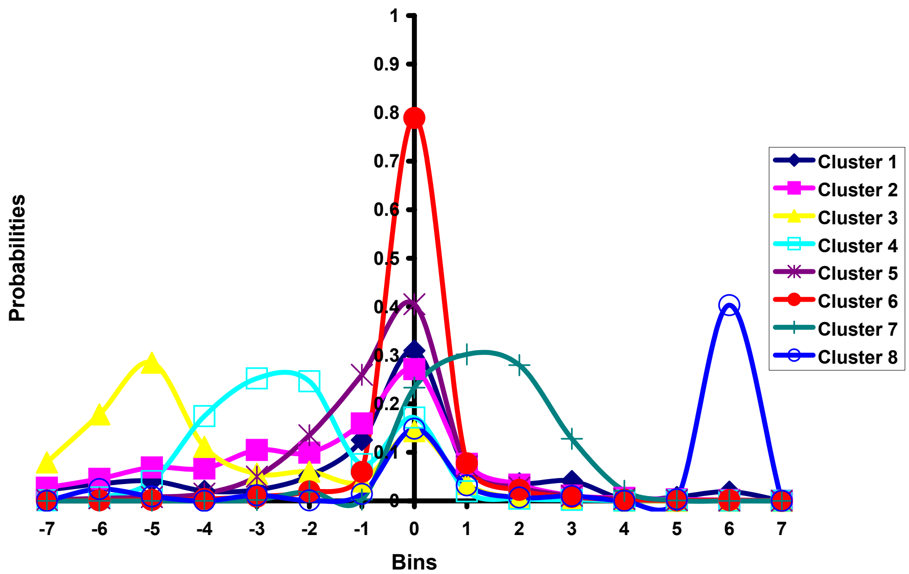


Figure 3.3: Average customer distributions for each cluster for K-Medoid Analysis.

centers. In our updated version of the heuristic we try to find cluster centers which minimize the maximum distance from any points in the cluster. Cluster centers are actual customers. The clustering technique is based on the assignment of each customer to the closest cluster center. The main issue in this clustering analysis is to map the customers according to a cluster center. By taking the sum of distances, customers can be well spread between the clusters. However, we try to look at the case when a customer has an irregular behavior and is assigned to a cluster whose center has similar types of behavior. In order to get a better spread among the clusters, we suggest Kohonen Networks in the next section.

In K-Medoid Analysis, we try to automate the process of recognizing irregular customer behavior. If a customer tends to have a forecast behavior which is different from the other customers, he may become a cluster center and pull other customers with similar behavior to his cluster. We do the clustering analysis for seven quarters of data for 2000 and 2001. Customers are assigned to eight different clusters. Each cluster represents a different customer behavior. In order to understand the properties of customers for each cluster, we plot the average customer distribution function for each cluster in Figure (3.5.2).

Cluster 6 represents the customers whose forecasts are true estimates of the actual orders about 80 percent of the time. Ideal customers also fall into cluster 6. Clusters 7 and 8 can be considered as over-estimators and Cluster 3 and 4 can be considered as under-estimators with different levels. Clusters 1, 2 and 5 show more irregular patterns that are mixes of over-, true or under-estimation.

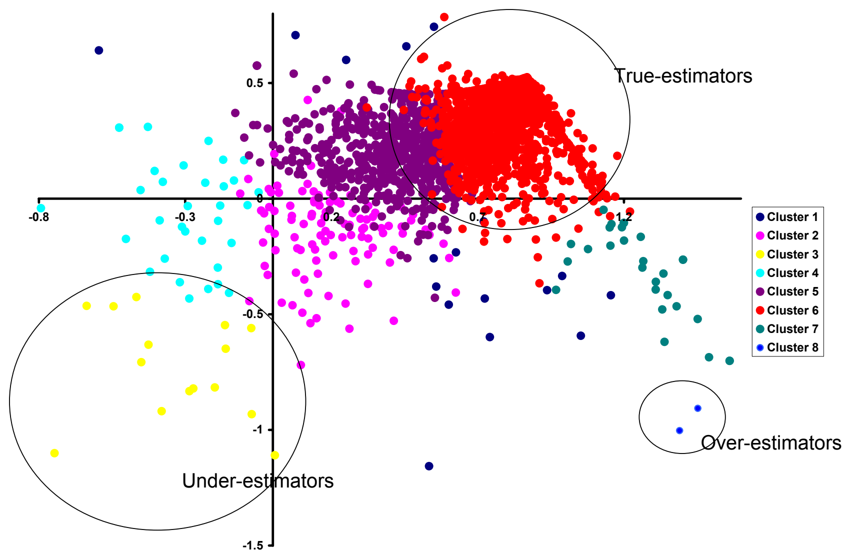


Figure 3.4: K-Medoid Analysis for years 2000 and 2001. Each point represents a customer.

In Figure (3.5.2), the Sammon's Mapping of the K-Medoid clustering is shown. Cluster 6 (red points), which we call the **ideal cluster**, includes the customers who are close to the ideal customer. As we get far from the ideal cluster, we can observe different clusters. Different regions on the mapping define different behaviors. The lower right part of the graph represents customers which are over-estimators (Clusters 7 and 8). Left of the y-axis represents the customers who are under-estimators. (Clusters 3 and 4). The other clusters (which have different behavioral patterns with different magnitudes) occupy different regions on the mapping.

In Figure (3.5.2), we observe the mapping for clusters for the top 11% of the customers who represent 80% of the total revenue. Most of the major customers fall into cluster 6

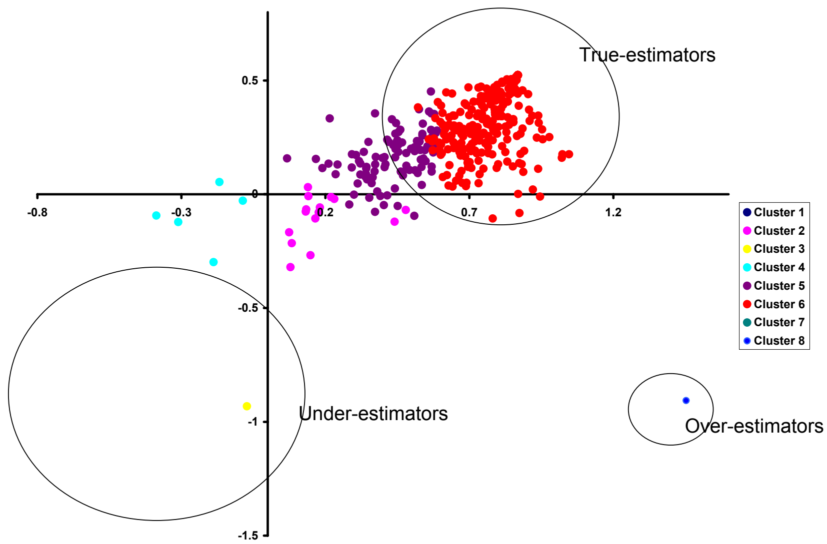


Figure 3.5: K-Medoid Analysis for years 2000 and 2001 for top 11% of the customers. Each point represents a customer.

which contains the ideal customer. We can still observe the other clusters with different behaviors. Therefore, our cluster analysis provides strong clusters.

Self-Organizing Maps (Kohonen Networks)

The Self-Organizing Maps (SOM) is an effective tool for the visualization of high-dimensional data. It converts the relationship of object in high-dimensional space into simple geometric relationship of their image points in a lower dimensional grid. SOM compresses information to display and produces some kind of abstraction. Kohonen (2001) described SOM as a nonlinear, ordered, smooth mapping of high-dimensional input data that manifolds onto the elements of a regular, low-dimensional array. SOMs are also named as Kohonen Networks.

We consider a SOM with one-dimensional eight clusters. Since SOMs are topological maps, distance between the Kohonen clusters represents the level of dissimilarity. The cluster numbers in K-Medoid Analysis do not provide any information about similarities of clusters. Compared to K-Medoid Analysis, Kohonen Network provides a better spread among the clusters. Figure (3.5.2) provides the average customer distributions for different clusters. Cluster 1 contains the customers who truly estimate their orders more than 90% of the time. When we compare ideal cluster of Kohonen Analysis (Cluster 1) with the ideal cluster of K-Medoid Analysis (Cluster 6), we observe that Kohonen Analysis has a small number of customers in ideal cluster with higher average true-estimation ratio than K-Medoid Analysis. Compared to K-Medoid Analysis, customer-quarter results between the clusters have more spread and it is harder to recognize irregular forecast behaviors compared to K-Medoid Analysis. To capture irregular patterns, K-Medoid Analysis seems to provide a better result. This is due to the fact that K-Medoid Analysis tries to minimize the maximum distance from cluster center. Therefore, any irregular customer-quarter result has to be close to a cluster center or become a cluster center himself in the output.

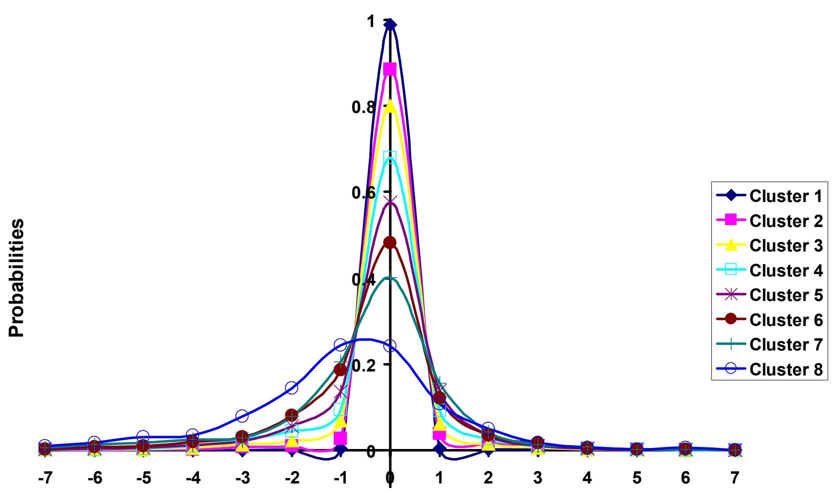


Figure 3.6: Average customer distributions for each cluster for Kohonen Network Analysis.

The Sammon's Mapping of Kohonen Analysis is a mountain like shape. As we get far from the ideal customer, we observe points which have similar distances from the ideal customer forming circular clusters. It is hard to recognize who are over-estimators or under-estimators. The advantage of Kohonen Analysis is having better spread and not giving too much weight to single observations. K-Medoid Analysis is better for recognizing outlier behaviors. When we analyze the major customers, we still observe all of the clusters. (Figure 3.5.2)

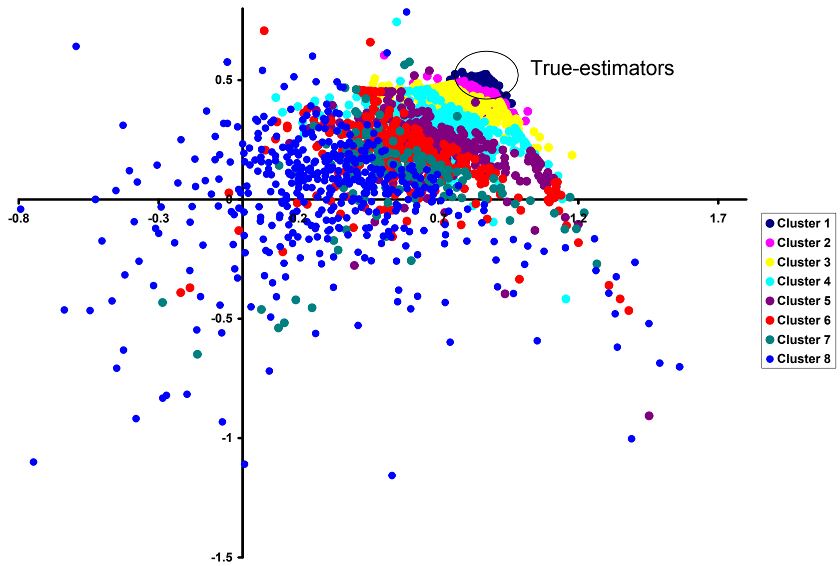


Figure 3.7: Kohonen Network Analysis for years 2000 and 2001. Each point represents a customer.

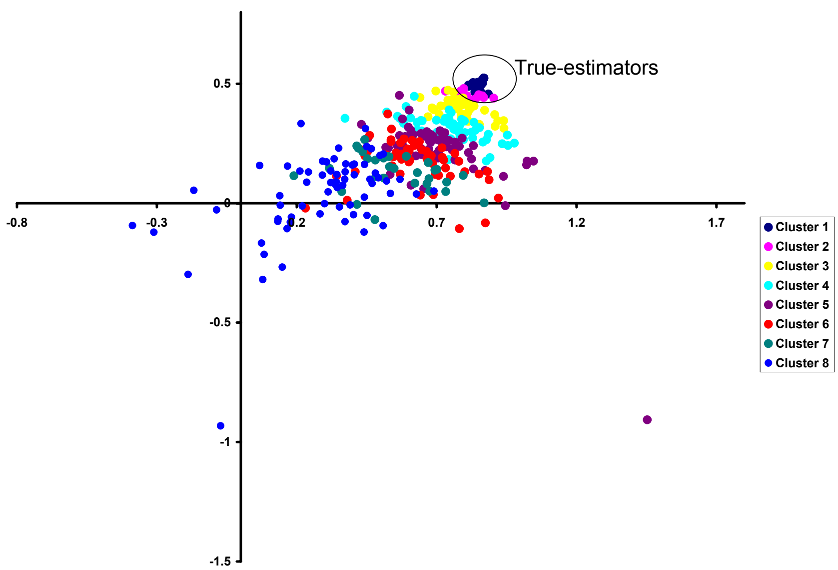


Figure 3.8: Kohonen Network Analysis for top 11% of the customers who represent 80% of the total revenue. Each point represents a customer quarter result.

Customers are Consistent:

In our analysis the clustering analysis provides significant clusters. In K-Medoid Analysis 36% stays in the same cluster for all quarters and 90% stays in the same cluster for at least half of the quarters. In Kohonen Analysis 15% stays in the same cluster for all quarters and 56% stays in the same cluster at least half of the quarters. Since some of the clusters in K-Medoid Analysis have large number of customers, these clusters occupy a larger volume of space. Therefore, the probability of staying in the same cluster is higher compared to Kohonen Analysis, where we have small clusters with equal size.

Reputation is Important:

In clustering analysis, we provide customer groups which provides irregular behaviors. When we take the averages of all customer distributions, we obtain a smoother behavior in Figure 3.5.2. Overall true-estimation (indexed 0) seems to be the most common behavior (67%). Under-estimation (indexed from -1 to -7) is a more common behavior (21%) compared to over-estimation (indexed from 1 to 7). The main reason for customers to under-estimate is to keep a good reputation with the supplier. In our problem environment, capacity is not a main concern for the supplier. The parts are highly customized and excess production is very costly for the supplier. Therefore, the cost of under-estimation is less compared to the cost of over-estimation for the supplier. In a long-term CPFR implementation this leads to the manufacturers provide forecasts which are going to be used with high confidence. In this way, excess production is minimized and the manufacturer builds a reputation with the supplier for accurate forecasts.

3.6 Managerial Use

Our customer forecast behavior analysis consists of many steps as discussed in the previous sections. Our procedure can be generalized as follows:

1. Orders are transformed into daily flows.

2. Accuracy ratios for daily forecasts are obtained by using the daily flows.

3. A quantitative representation of each customer's forecast is computed by customer bin analysis.

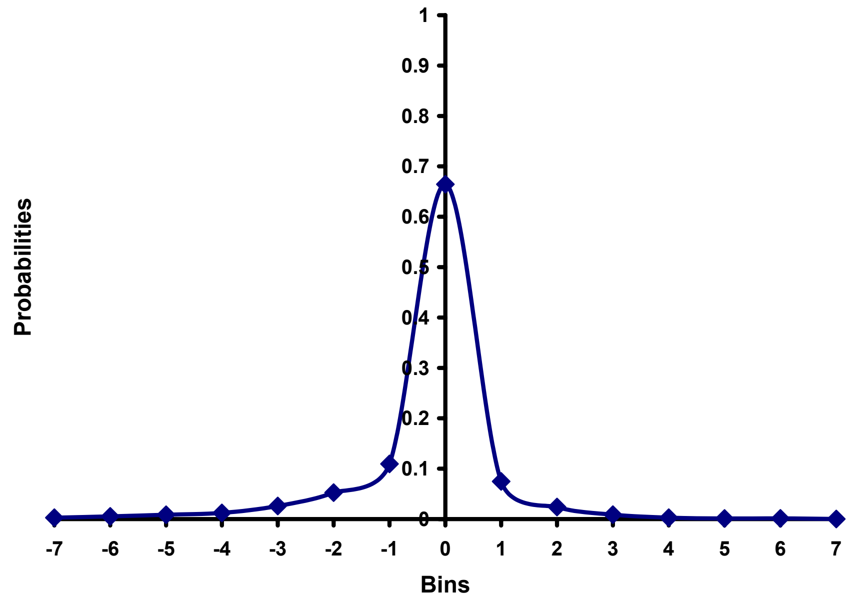


Figure 3.9: The average distribution for all customers

4. Customer order distributions are used as inputs to clustering.

5. The clusters are projected on a two-dimensional graph by using Sammon's Mapping in order to visualize the clusters.

Our customer forecast behavior analysis has several useful applications for the practitioners.

Automation for Detecting Irregular Customer Behavior: When the data set is too large, it is hard to recognize irregularities in a particular customer's forecast. Signalling mechanisms can be equipped to give quick responses. For example, a group of customers can start to give aggressive orders, which directly affect the capacity decisions of the supplier. An automated signaling mechanism can be designed with the help of K-Medoid Analysis. K-Medoid Analysis is based on assigning each customer to the closest cluster center. Therefore, when a customer behaves differently from the other customers, he either initiates a new cluster or joins the cluster of other customers with similar behaviors. Therefore, this gives a signal to the supplier to take necessary actions.

Creating Performance Benchmarks: The forecasting performance of a customer is dependent upon many outside factors such as trends in the industry, competition, exchange rates, etc. It is hard to quantify the effect of each factor on a customer's forecast performance. Therefore, the supplier needs to design flexible performance benchmarks which quickly and dynamically adapt to changes without any supervision. Our clustering analysis discussed in Section 3.5.2 forms customer groups with similar forecast performances. The cluster with high forecast accuracy represents the ideal customer behavior. As discussed in Section 3.5.2, Kohonen Network is a clustering technique which provides a good spread of customers to the clusters. By performing Kohonen Analysis for every quarter, the supplier can dynamically update the properties of an ideal customer and report this to the customers as performance benchmarks. Kohonen clusters are topological maps and the distance between the Kohonen clusters represent the level of dissimilarity. In Figure (3.5.2) from cluster 1 to cluster 8, the percentage of customers who provide true forecasts decrease. Therefore, as the cluster number increases, the magnitude of the deviation from the ideal behavior increases.

Group Monitoring Options: A customer's forecast behavior can be affected by other customers' actions. Therefore, understanding behavior of related customers is important in decision-making. Customer groups can be formed for close monitoring. For example, in the introduction of a new item to the market, customers are in the initial phases of learning the market conditions. A separate analysis can be performed on this group of customers who are ordering the item. For the items with limited supplier capacity we can observe aggressive ordering from the customers in order to get more of the supplier's capacity.

Quarterly Reports: Quarterly reports can be generated by the supplier based upon the forecast performance of the customers. Customers can be informed about their performance and make improvements in their forecasting process. Our mapping analysis can be used at that point in order to provide the position of the customer with respect to the other customers. Reward/penalty schemes can be developed based on the performance reports of the customer. The trajectory of a customer for different quarters can be interpreted with respect to different points on the mapping (such as the position of ideal customer, his previous quarter position or other customers' or competitors' performances).

Customer-Level Analysis: Our analysis can be repeated for different parts of the same supplier. By doing that, the supplier can understand if the customer is consistent with his behavior for all the parts he is ordering. Parts can be an effective factor on the performance of a customer. A customer can be an over-estimator for one part. However, the same customer can turn out to be a true-estimator for other parts.

3.7 Conclusion and Future Research

In our analysis we describe complexities such as order-splitting, order-combining and shifts in due dates. We disaggregate the orders into daily flow analysis to overcome complexities and to compute the accuracy ratios of the forecasts. Another solution for handling the complexities is aggregating the orders. In our recent analysis with year 2002 and 2003 data, we search for the minimum aggregation level that eliminates the complexities with the orders. The major disadvantage of the aggregation approach is the loss if information. However, aggregation transforms the data into a format that is compatible with standard statistical parametric techniques.

We employ customer bin analysis to obtain customer order distributions. This vector has been used in our analysis as a quantitative representation of a customer's performance. Using the vectors obtained from customer bin analysis, we categorize customer order distributions into data clusters and use projection techniques for visualization. We derive a distance metric compatible with our problem environment in order to compute the similarities among the customers.

In our analysis, we show that customers are consistent with their forecast behavior. Some customers consistently provide bad forecast performance. The supplier should take the necessary actions to improve the forecasts of these customers. However, there are several supply chain features that complicate handling customer forecasts. Forecast effort is a hidden action of the customer. It may be costly for the customer to spend time and effort to provide accurate forecasts. Decreasing forecast effort increases the forecast variability. Since the forecast action is not verifiable, it is also not contractable. The supplier can overcome problems with the forecast variability by rewarding customers based upon observable outcomes (their position in clustering and mapping).

When a customer has a conflict of interest with the supplier, he can provide biased forecasts by inflating or deflating his order size. Over-estimation causes excess production for the supplier and under-estimation leads to lost sales. When the sum of customer orders exceeds the suppliers' capacity, capacity rationing mechanisms lead customers to over-estimate the orders. However, we observe that under-estimation is a more common behavior in our analysis. Therefore, capacity is not the main concern of the customers while placing orders. Reputation concerns become more critical and cause customers to under-estimate. Since the relationships are long term in the automotive industry, customers (auto manufacturers) tend to keep good relationships with the supplier. The analysis of dynamic models with reputation effect can provide insights into eliminating the bias in customers' forecasts.

Chapter 4

Using Customers' Reported Forecasts to Predict Future Sales

In the last two decades, business environments have been undergoing rapid changes due to expansion of computer resources. Companies can now share huge amounts of information across their supply chains. Empowered with this information companies have a large spectrum of choices to create complex products and processes. Due to increasing complexity, we observe more collaborative effort between the different parties in the supply chain. To achieve higher returns, suppliers and customers invest in technologies that provide real-time access to demand, inventory, price, sourcing, and production data. Sharing information is key to increasing the profits under demand uncertainty. A critical factor that determines the quality of information transmission is its reliability. Inaccurate information can lead to severe costs for parties in the supply chain. Most of the existing methods are not appropriate because of the potential bias in the forecasts. However, the suppliers still collect forecasts from the customers and do not only use their own forecasts. The biased forecasts still contain information which might be useful for the supplier. It is critical for the supplier to detect the bias and predict the future order of a customer.

We provide an empirical study about forecast sharing in the supply chains. We analyze the forecasts of final orders that are received by an automotive supplier who produces multiple parts for auto manufacturers. Forecast information helps the supplier to predict the final orders to do production planning in advance. As part of the collaborative effort, customers provide forecast updates to the supplier at different order dates.

Example 10 *Forecasts updates for 1-Aug, 8-Aug and 15-Aug*

1-Aug		**8-Aug**		**15-Aug**	
Due Date	**Quantity**	**Due Date**	**Quantity**	**Due Date**	**Quantity**
8-Aug	*800*				
15-Aug	*800*	*15-Aug*	*160*		
22-Aug	*960*	*22-Aug*	*480*	*22-Aug*	*320*
29-Aug	*960*	*29-Aug*	*640*	*29-Aug*	*640*
5-Sep	*960*	*5-Sep*	*800*	*5-Sep*	*800*
		12-Sep	*960*	*12-Sep*	*800*
				19-Sep	*960*

Example 10 represents a typical forecast for a customer. The customer provides different updates in August 1^{st}, August 8^{th} and August 15^{th}. For example in forecast date August 1^{st}, the customer provides forecasts for the next five weeks with quantities 800, 800, 800, 960 and 960. In each forecast date, the customer updates the forecasts from the previous forecast date and can place new orders. Therefore, the forecasts can be considered as a flow of orders which evolve over time. As can be noticed from the forecast values due to production and transportation constraints, the customer provides forecasts which are multiples of some lot sizes. Therefore, we can divide all the forecasts by a common divisor (160) to obtain the number of batches in each order. In this analysis, we propose a framework for modeling the forecast generation process at the customer.

Forecasts provide information about the future orders of the customer. However, the forecast information can be quite noisy and can be misleading for the supplier. It is critical for the supplier to process the forecasts to estimate future orders. In Chapter 3, we study a similar data set in a non-parametric framework and provide empirical support for downstream players consistently over- or under-estimating their forecasts through time. It is important for a supplier to recognize a significant pattern in the forecasts by looking at order history of a customer. Our objective here is to provide a framework to extract information from the forecast data and adjust the forecasts to provide a better estimate of future orders. For example, if a customer constantly overestimates his orders, then the supplier can detect this behavior and can remove the bias from the forecast.

Why do customers provide poor forecast performance?

There are a couple factors that can lead to forecast errors.

1. **Uncertainty in the usage:** There is always uncertainty in the customer's system due to demand variance, lead times, machine failures, etc. Therefore, the error in the forecast can be a result of these factors. As the due date approaches, the customer has more information about the demand and uncertainty decreases.

2. **Bias in the forecast:** The customer can have different costs for overestimation and underestimation. When the customer overestimates, the supplier can penalize the customer or the customer can lose the goodwill of the supplier. In the case of underestimation, the customer cannot satisfy the demand, which can lead to delays in production and potentially lost sales. The customer must consider the trade off between overestimation and underestimation and submit a forecast which minimizes the expected cost. The unit overestimation cost is not necessarily equal to underestimation cost, therefore the customer may add a bias to his forecast to minimize his cost.

Therefore a realistic model should consider the uncertainty in production and bias in the forecast.

4.0.1 Literature Review

Forecasting has been addressed in many different problem settings. There is huge body of literature in forecasting new observations. In our analysis, the suppliers collect self-reported forecasts which might be biased from the customers. We use a bayesian approach to estimate the model parameters in our analysis. Geweke and Whiteman (2006) provide an extensive review of literature in bayesian forecasting models. Another stream of research which deals with self-reported forecasts is consensus forecasting models. Batchelor and Dua (1995)) show that combinations of different forecasts even from a small number of sources is helpful in predicting future values. In our case, we have regular updates from the customers which are combined to predict the customer demand. Therefore, the final prediction incorporates different updates of the customer.

In our analysis, we model the demand function as a time-series process. The customer submits his final forecast after considering the cost of overestimation and underestimation. Time-series assumption has been also used by some other researchers to provide theoretical results about forecasting. Graves (1999) assumes a non-stationary demand with ARIMA process and shows that inventory decisions behave much differently compared to a stationary process. Aviv (2003) proposes a unified time-series framework for forecasting and inventory control. He assumes that different parties observe different subsets of information and adopts their forecasting and demand process accordingly. Aviv (2001, 2002) also study models with time-series demand. Our results provide an empirical support for forecasting models which assume strategic behavior of parties with private information in a time-series framework. Chen et al. (2000a, 2000b) assumes that the downstream player uses moving average forecasts to place orders to a supplier. They measure the amplification in the variance of the orders which is known as bullwhip effect.

There are other mathematical models in literature for the evolution of demand. Graves et al. (1986a, 1986b, 1998) and Heath and Jackson (1994) develop the Martingale Model of Forecast Evolution (MMFE) to model the evolution of forecasts. In MMFE, a forecaster creates forecasts for the planning horizon and updates them in regular intervals. The error of the forecast updates are assumed to follow the Martingale Property: independent, identically distributed, multivariate normal random variables with mean 0. MMFE approach has been studied by a number of researchers under different problem settings. (see, e.g., Gullu (1996) , Graves et al. (1998), and Toktay and Wein (2001)). Another approach is considering that some demand parameters are unknown in advance and using Bayesian updates to incorporate new information as it becomes available (Scarf (1959) , Azoury (1985) and Lariviere and Porteus (1999))

Our analysis has several aspects that have not been addressed in the literature using self reported forecasts. First, we assume a loss function for the forecast errors and explain the dynamics behind the loss function by using a time-series demand. Our analysis is the first to put the ARIMA form into a newsvendor framework with a cost of overestimation and underestimation. Second, we assume forecast values which are discrete values since they are multiples of lot sizes. Third, we provide a statistical procedure to estimate the model parameters for this complicated problem. A limitation is that we will not consider a symmetric game where customer adjusts to the suppliers and vice versa.

The empirical research about forecasting is very limited in the supply chain literature. Terwiesch et al. (2003) considers the problem from a buyer's perspective. He demonstrates that poor forecast performance, in terms of forecast inflation and volatility, damages the buyer's reputation and leads to a lower service. In our analysis, we looked at the problem from the supplier's perspective and provide analysis to understand the forecast behavior of the customers.

4.1 Problem Environment

In our analysis, we model the orders that are placed to an automotive parts supplier by auto customers. Customers place some preliminary orders (forecasts) starting from six months before the order date and adjust the forecasts before the due date. The parts are engine systems and generate multi-billion dollar revenue for the supplier. The customer has a better ability to predict the demand due to its proximity to the final demand. The supplier can only observe the forecasts submitted by the customer. In Figure 4.1, we can see that the manufacturers provide different updates at 1-Aug, 8-Aug and 15-Aug for the next 5 weeks.

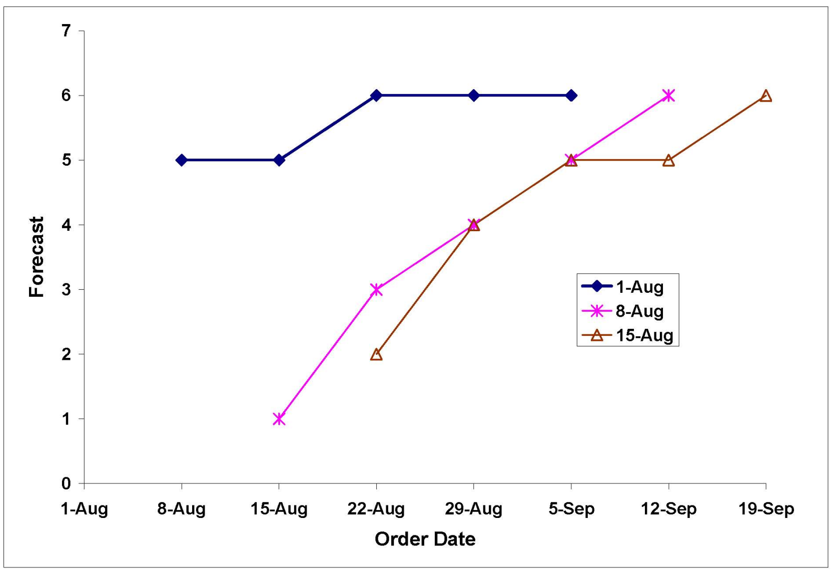


Figure 4.1: The forecast updates at different order dates 1-Aug, 8-Aug and 15-Aug.

4.1.1 Demand Model

In our analysis, we observe that the forecasts and forecast errors are autocorrelated through time. A forecast error leads to excess inventory or backordering and carries to the next period. Another important factor is the autocorrelation between the demand values of consecutive periods. For example, high demand periods can be followed with low demand. Therefore, the demand model should be fairly adaptive in order to incorporate the available information in each period. ARMA model provides a flexible model to describe different demand processes. (Box and Jenkins 1970) Let X_t as the demand of the customer at time

t. We model X_t as

$$X_t - \mu = \phi(X_{t-1} - \mu) + \epsilon_t - \theta\epsilon_{t-1} \tag{4.1}$$

$$X_t = (1 - \phi)\mu + \phi X_{t-1} + \epsilon_t - \theta\epsilon_{t-1} \tag{4.2}$$

which is $ARMA(1,1)$ with $E(X_t) = \mu$. Having both autoregressive and moving average components, the $ARMA(1,1)$ model can capture the significant behaviors in the data. Our framework can be extended over higher number of lags. We assume that the error ϵ_t is normally distributed with $E(\epsilon_t) = 0$ and $Var(\epsilon_t) = \sigma^2$.

4.1.2 Forecast Generation Model

We assume that forecasts are generated in a three stage process:

i. At time t, the customer considers the next K periods based on his available information set Ω_t. He computes the distribution $F(X_{t+1}, ..., X_{t+K}|\Omega_t)$ for the next K periods.

ii. The customer derives his optimal forecasts $\bar{X}_t = (\bar{X}_t(1), ..., \bar{X}_t(K))$ using an asymmetric loss function which represents the strategic behavior of the customer.

iii. At time t, the customer submits a forecast vector to the supplier, $\hat{X}_t = (\hat{X}_t(1), ..., \hat{X}_t(K))$ where $\hat{X}_t(k) = \bar{X}_t(k) + \gamma_t(k)$ for $k = 1, ..., K$. Errors $\gamma_t(k)$ might perturb the process to generate non-ARIMA forecasts.

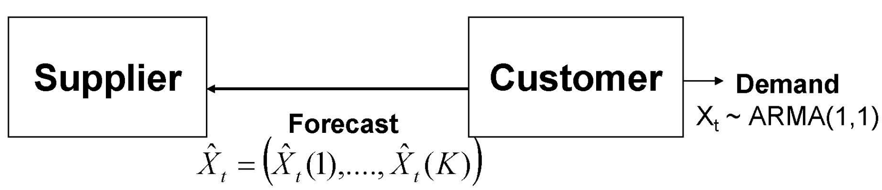


Figure 4.2: Manufacture first observes demand and submit a forecast to the supplier

The customer observes demand which takes continuous values. However, from Example 1, we observe that the customer places integral forecasts which are multiples of lot sizes.

Therefore, we assume that the forecasts can only take integer values. Next we explain each step of the forecast generation model in detail.

Derivation of $F(X_{t+1}, ..., X_{t+K}|\Omega_t)$ at time t

By using $ARMA(1,1)$, we also model the uncertainty in demand for future periods. At time t, the customer can derive the k-step ahead demand X_{t+k} as

$$X_{t+1} = (1-\phi)\mu + \phi X_t + \epsilon_{t+1} - \theta\epsilon_t \tag{4.3}$$

for $k = 1$ and

$$X_{t+k} = (1-\phi^k)\mu + \phi^{k-1}(\phi X_t - \theta\epsilon_t) + \sum_{n=1}^{k-1} \epsilon_{t+n}\phi^{k-n-1}(\phi-\theta) + \epsilon_{t+k} \tag{4.4}$$

for $k = 2, ..., K$. Since the unobserved values $\epsilon_{t+1}, ..., \epsilon_{t+k}$ are Normal$(0, \sigma^2)$, we can also show that X_{t+k} is normally distributed with

$$\begin{aligned} E(X_{t+k}|\Omega_t) &= (1-\phi^k)\mu + \phi^{k-1}\lambda_t \quad & k = 1, ..., K \\ Var(X_{t+k}|\Omega_t) &= \sigma^2 \upsilon_k^2 \quad & k = 1, ..., K \end{aligned}$$

where $\lambda_t = (\phi X_t - \theta\epsilon_t)$, $\upsilon_1 = 1$ and $\upsilon_k = \sqrt{(1 + (\phi-\theta)^2\left(\frac{1+\phi^{2(k-1)}}{1-\phi^2}\right)}$).

Loss function

The customer may have private information which is not available to the supplier. The customer provides forecasts to minimizes his own cost. We model this strategic behavior as an asymmetric loss function with different values for costs of overestimation and underestimation. Therefore, the customer does not necessarily submit the expected demand as his forecast and can add a bias to his forecast. We provide an empirical framework to test the significance of this hypothesis by looking at the forecast data.

The customer can evaluate the cost of overestimation and underestimation to adjust his final forecast. A common assumption in supply chain management literature is having a linear cost function with c_o for each unit of overestimation and c_u for each unit of underestimation (Cachon 2004). These costs can be considered as the imputed cost of forecast errors. Cost of overestimation can include the penalty cost and goodwill of the supplier. The cost of underestimation can be composed of the expected backordering cost or lost sales. By following the same approach, we can write the loss function as

$$U(X_t(k), X_{t+k}) = \begin{cases} c_o|X_t(k) - X_{t+k}| & \text{if } X_t(k) \geq X_{t+k} \\ c_u|X_{t+k} - X_t(k)| & \text{otherwise} \end{cases}$$

where $c_o, c_u > 0$.

The customer can derive the expected loss (or risk) for any forecast $X_t(k)$ as follows

$$E(U(X_t(k), X_{t+k})|\Omega_t) = c_o \int_{-\infty}^{X_t(k)} (X_t(k) - X_{t+k})\, dF(X_{t+k}) + c_u \int_{X_t(k)}^{X_t=\infty} (X_{t+k} - X_t(k))\, dF(X_{t+k})$$

Submission of the Forecast

Therefore, at period t, the customer solves a minimization problem to minimize his expected loss function $E(U(X_t(k), X_{t+k}))$ and finds an optimal forecast

$$\bar{X}_t(k) = \text{argmin } E(U(X_t(k), X_{t+k})). \tag{4.5}$$

If we assume continuous valued forecasts, we can take the first derivative of the expected loss with respect to $X_t(k)$ to find the optimal continuous valued forecast $\zeta_t(k)$,

$$\begin{aligned} \frac{\partial E(U(X_t(k), X_{t+k})|\Omega_t)}{\partial X_t(k)} &= c_o \int_{-\infty}^{X_t(k)} X_{t+k} dF(X_{t+k}) - c_u \int_{X_t(k)}^{X_t=\infty} X_{t+k} dF(X_{t+k}) \\ &= c_o F(X_t(k)|\Omega_t) - c_u\,(1 - F(X_t(k)|\Omega_t)) \\ &= F(X_t(k)|\Omega_t)(c_o + c_u) - c_u \end{aligned}$$

Since $\frac{\partial^2 E(U(X_t(k),X_{t+k})|\Omega_t)}{\partial X_t(k)^2} = f(X_t(k)|\Omega_t) > 0$, from the first order conditions $\zeta_t(k)$ satisfies,

$$\begin{aligned} F(\zeta_t(k)|\Omega_t)(c_o + c_u) - c_u &= 0 \\ F(\zeta_t(k)|\Omega_t) &= \frac{c_u}{c_o + c_u} \end{aligned}$$

Therefore, $\zeta_t(k)$ for the customer is

$$\zeta_t(k) = F^{-1}\left(\frac{c_u}{c_o + c_u}|\Omega_t\right). \tag{4.6}$$

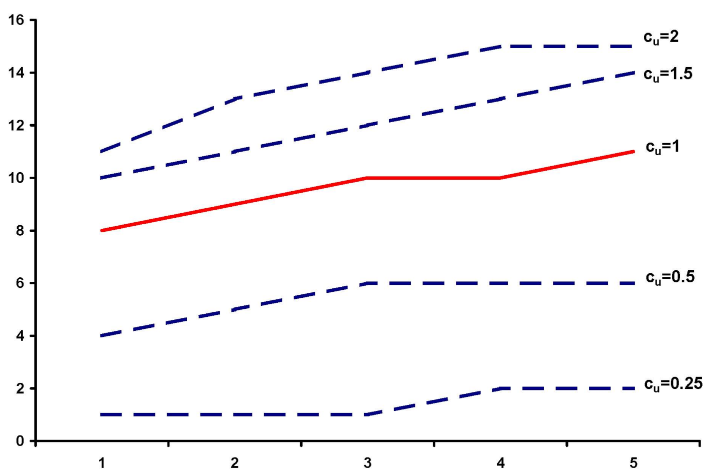


Figure 4.3: The optimal k-step ahead forecasts at time t for $\mu = 12$, $\sigma = 8$, $\phi = 0.7$, $\theta = 0.1$, $X_t = 6$, $\epsilon_t = 3$. The solid line represents an unbiased forecast with $c_u = c_o = 1$.

When the loss function is symmetric ($c_u = c_o$), the customer finds the median $F^{-1}(0.5|\Omega_t)$ as the optimal forecast. If the distribution is symmetric, then median is equal to mean and the customer submits $E(X_{t+k}|\Omega_t)$ as the optimal forecast. As can be seen in Figure 4.3 the customer adds a bias to his forecast when the the loss function is asymmetric. When $c_u > c_o$, the customer tends to overestimate his orders and adds a positive bias to his forecast. When $c_u < c_o$, it is less costly to underestimate for the customer. In this case the customer has a negative bias in his forecast. Therefore, asymmetric loss functions can explain the dynamics behind the bias in the forecasts.

When we assume the the ARMA(1,1) model, the k-period ahead optimal continuous valued forecast can be derived as :

$$\zeta_t(k) \quad = \quad (1-\phi^k)\mu + \phi^{k-1}\lambda_t + \tau\sigma\upsilon_k \quad k = 1, ..., K$$

where τ be the z-value of $\frac{c_u}{c_o+c_u}$ for standard normal distribution.

In our analysis, we assume that forecasts can only take integer values. Since $E(U(X_t(k), X_t(k)))$ is convex, the optimal integer valued forecast $\bar{X}_t(k)$ is

$$\bar{X}_t(k) = \begin{cases} \lfloor \zeta_t(k) \rfloor & \text{if } E(U(\lfloor \zeta_t(k) \rfloor, X_t(k))) < E(U(\lceil \zeta_t(k) \rceil, X_t(k))) \\ \lceil \zeta_t(k) \rceil & \text{otherwise} \end{cases}$$

where $\lfloor \cdot \rfloor$ is the floor function and $\lceil \cdot \rceil$ is the ceiling function.

We assume that there might be some errors $\gamma_t(k)$ added to the optimal integer valued forecasts to generate non-ARIMA forecasts. The final forecast is

$$\hat{X}_t(k) = \bar{X}_t(k) + \gamma_t(k)$$

where $\gamma_t(k)$'s are random integral errors. We assume that the forecasts in the errors are $\gamma_t(k)$ generated by a mixture of two discrete normal distributions

$$\gamma_t(k) \;\sim\; \begin{cases} N(0, \omega_1) & \text{with probability p} \\ N(0, \omega_2) & \text{with probability 1-p} \end{cases}$$

where p is the mixing probability.

The matrix form of the problem is available in Appendix 5.2.2.

4.2 Supplier's Problem

At $t = T$, the supplier observes the forecast matrix $\hat{X}$ which consists of the forecast vectors $\left\{\hat{X}_1, ..., \hat{X}_T\right\}$ that are submitted at $t = 1, .., T$. By following the forecast generation model, at time t the customer submits a forecast based on the demand parameters (μ, ϕ, θ and σ), cost parameters (c_u and c_o) and recent information (X_t and ϵ_t). For the sake of simplicity we will denote the parameters as $\Lambda = (\mu, \phi, \theta, \sigma, c_u, c_o, X_0, \epsilon_0)$. When Λ and previous demand values X is available, it is straightforward to determine the forecast. However supplier cannot observe Λ and X. The supplier can only make inference about the demand and cost information of the customer by only looking at the forecast. Therefore, supplier's problem can be formulated as finding the distribution $P(\Lambda|\hat{X})$. The problem has the following complexities:

1. The final demand of the customer is assumed to be continuous. However, the customer provides integer forecasts which are multiples of production lot sizes.

2. The customer does not necessarily have a symmetric loss function. The customer's submits biased forecasts when $c_u \neq c_o$.

3. The supplier does not observe demand $X = (X_1, ..., X_T)$, nor the demand and cost parameters(Λ).

4. The forecasts can come from non-ARIMA process due to errors $\gamma_t(k)$. The supplier does not know $\Upsilon = (p, \omega_1, \omega_2)$ for $\gamma_t(k)$.

The supplier seems to have very limited information once we consider the complexity of the analysis. However, we provide a Bayesian model which can be used to make inference about the model parameters. Forecast vectors provide substantial amount information about the forecast generation model of a customer. At time t, the supplier observes forecasts for each of the following K periods. Therefore, the supplier has access to TK data points for T periods.

By using a Bayesian framework we first define priors for the model parameters:

$$\begin{vmatrix} \mu \\ \phi \\ \theta \\ \log(\sigma) \\ \log(c_u) \end{vmatrix} \sim N(\hat{\Lambda}, V)$$

$$\epsilon_0 \sim N(0,\sigma) \quad X_0 \sim N(\mu, \sigma\sqrt{\tfrac{1+\theta^2-2\theta\phi}{1-\phi^2}}) \quad p \sim Beta(a,b) \quad log(\omega_i) \sim N(0, v_i) \text{ for i=1,2}$$

Since the ratio of $\frac{c_u}{c_u+c_o}$ is important, we assume that $c_o = 1$.

Although we define priors for the parameters, we cannot still derive the distribution of $P(\Lambda|\hat{X})$. Therefore, we use gibbs sampling in order to make draws from Λ by defining X as a latent parameter. We iteratively sample

1. $\Lambda|X, \bar{X}$

2. $X, \bar{X}|\Lambda, \hat{X}, \Upsilon$

3. $\Upsilon|\bar{X}, \hat{X}$.

By using Bayes Formula, we can write the first step where we draw $\Lambda|X, \bar{X}$ as

$$P(\Lambda|X, \bar{X}) = \frac{P(\Lambda, X, \bar{X})}{P(X, \bar{X})} = \frac{P(X, \bar{X}|\Lambda)P(\Lambda)}{P(X, \bar{X})} \tag{4.7}$$

Since $P(X, \bar{X})$ is constant we have

$$P(\Lambda|X, \bar{X}) \quad \propto \quad L(\Lambda; X, \bar{X}) \cdot P(\Lambda) \tag{4.8}$$

where $L(\Lambda; X, \hat{X})$ is the likelihood of Λ for any given values of X and $\hat{X}$. $L(\Lambda; X, \hat{X})$ is zero if X and λ does not generate $\bar{X}$ as the optimal forecast. Therefore, the likelihood can be written as

$$L(\Lambda; X, \bar{X}) = \begin{cases} L(\Lambda; X) & \text{if } \Lambda \text{ and } X \text{ generates } \bar{X} \\ 0 & \text{otherwise} \end{cases}$$

where

$$L(\Lambda; X) = \left(\frac{1}{2\Pi\sigma^2}\right) exp\left\{-\frac{1}{2\sigma^2}\sum_{t=1}^{T}\epsilon_t^2\right\} \tag{4.9}$$

4.3 Estimation

The supplier wants to derive the distribution of parameters given the forecast matrix, $P(\Lambda|\hat{X})$. Our sampler is a modified slice sampler with rejection. The general slice sampling algorithm (Neal 2003) is constructed using the principle that one can sample from a distribution by sampling uniformly from the region under the plot of its density function.

4.3.1 Mechanism of the Slice Sampler

Assume that we want to sample from a distribution for a variable $x \in R^n$, which has density function $f(x)$. We can introduce an auxiliary variable real variable, y and define joint distribution function $p(x, y)$ of x and y which is uniformly distributed over the region $S = \{(x, y) : 0 < y \leq f(x)\}$. S is the area under $f(x)$. Let $Z = \int f(x)$, then we have

$$p(x, y) = \begin{cases} 1/Z, & \text{if } 0 < y < f(x) \\ 0, & \text{otherwise} \end{cases}$$

The marginal density for x is

$$p(x) = \int_0^{f(x)} 1/Z dy = f(x)/Z \tag{4.10}$$

We can sample jointly for (x, y) and keep x values to replicate $f(x)$.

4.3.2 Our Sampler for Estimation

We first find an initial set of Λ, X, $\bar{X}$ and υ which can generate the forecast matrix $\hat{X}$. We then iteratively sample

1. $\Lambda | X, \bar{X}$
2. $X, \bar{X} | \Lambda, \hat{X}, \Upsilon$
3. $\Upsilon | \bar{X}, \hat{X}$.

Our sampler works as follows:

Iteration 0 (Initialization):

1. Draw $\mu^{(0)}$,$\phi^{(0)}$, $\theta^{(0)}$, $\sigma^{(0)}$, $c_u^{(0)}$, $\epsilon_0^{(0)}$ and $X_0^{(0)}$ from the prior distributions.

$$\left| \begin{array}{c} \mu^{(0)} \\ \phi^{(0)} \\ \theta^{(0)} \\ \log(\sigma^{(0)}) \\ \log(c_u^{(0)}) \end{array} \right| \sim N(\hat{\Lambda}, V)$$

$$\epsilon_0^{(0)} \sim N(0, \sigma) \quad X_0^{(0)} \sim N(\mu, \sigma\sqrt{\tfrac{1+\theta^2-2\theta\phi}{1-\phi^2}})$$

where $-1 \leq \phi^{(0)} \leq 1$ and $-1 \leq \theta^{(0)} \leq 1$.

2. Draw $\epsilon^{(0)} = \left\{\epsilon_1^{(0)}, ..., \epsilon_T^{(0)}\right\}$ from $N(0, I\sigma^2)$. Compute $X^{(0)}$ and optimal $\bar{X}^{(0)}$.

$$X_t^{(0)} = (1 - \phi^{(0)})\mu^{(0)} + \phi^{(0)} X_{t-1}^{(0)} + \epsilon_t^{(0)} - \theta\epsilon_{t-1}^{(0)}$$

and

$$\bar{X}_t(k)^{(0)} = \text{argmin } E(U(X_t(k), X_{t+k})). \tag{4.11}$$

for $t = 1, ..., T$ and $k = 1, ..., K$.

3. Draw $p^{(0)}$, $\omega_1^{(0)}$ and $\omega_2^{(0)}$ from the priors

$$p^{(0)} \sim Beta(a,b) \quad log(\omega_1^{(0)}) \sim N(0,v_1) \quad log(\omega_2^{(0)}) \sim N(0,v_2)$$

Repeat the following for a specified number (I) of iterations. (for i=1,...,I)

Iteration i:

1. In this step, we generate $\Lambda^i|X^{(i-1)}, \bar{X}^{(i-1)}$. So we draw

   - $\mu^{(i)}|\phi^{(i-1)}, \theta^{(i-1)}, c_u^{(i-1)}, \sigma^{(i-1)}, \epsilon_0^{(i-1)}, X_0^{(i-1)}, X^{(i-1)}, \bar{X}^{(i-1)}$
   - $\phi^{(i)}|\mu^{(i)}, \theta^{(i-1)}, c_u^{(i-1)}, \sigma^{(i-1)}, \epsilon_0^{(i-1)}, X_0^{(i-1)}, X^{(i-1)}, \bar{X}^{(i-1)}$
   - $\theta^{(i)}|\mu^{(i)}, \phi^{(i)}, c_u^{(i-1)}, \sigma^{(i-1)}, \epsilon_0^{(i-1)}, X_0^{(i-1)}, X^{(i-1)}, \bar{X}^{(i-1)}$
   - $c_u^{(i)}|\mu^{(i)}, \phi^{(i)}, \theta^{(i)}, \sigma^{(i-1)}, \epsilon_0^{(i-1)}, X_0^{(i-1)}, X^{(i-1)}, \bar{X}^{(i-1)}$
   - $\sigma^{(i)}|\mu^{(i)}, \phi^{(i)}, \theta^{(i)}, c_u^{(i)}\epsilon_0^{(i-1)}, X_0^{(i-1)}, X^{(i-1)}, \bar{X}^{(i-1)}$
   - $\epsilon_0^{(i)}|\mu^{(i)}, \phi^{(i)}, \theta^{(i)}, c_u^{(i)}, \sigma^{(i)}X_0^{(i-1)}, X^{(i-1)}, \bar{X}^{(i-1)}$
   - $X_0^{(i)}|\mu^{(i)}, \phi^{(i)}, \theta^{(i)}, c_u^{(i)}, \sigma^{(i)}, \epsilon_0^{(i)}, X^{(i-1)}, \bar{X}^{(i-1)}$

   by using a *sequential slice sampler*. Here, we show how we draw $\theta^{(i)}$. The same analysis is repeated for all the above parameters in the given order.

   (a) Since we always guarantee to have feasibility of $\bar{X}$ in each step, we can compute the likelihood

   $$\begin{aligned} L_\theta^{(i-1)} &= L(\Lambda; X^{(i-1)}, \bar{X}^{(i-1)}) \\ &= L(\Lambda; X^{(i-1)}) \\ &= P(X_1^{(i-1)}, ..., X_T^{(i-1)}|\mu^{(i)}, \phi^{(i)}, \theta^{(i-1)}, c_u^{(i-1)}\epsilon_0^{(i-1)}, X_0^{(i-1)}) \end{aligned}$$

   This is our vertical slice.

   (b) Draw a random variable u_θ from Uniform$(0, L_\theta^{(i-1)})$. This is our horizontal slice.

   (c) While true,

      i. Draw $\theta^{(i)}$ from the prior distribution $\theta^{(i)}|\mu^{(i)}, \phi^{(i)}, c_u^{(i-1)}, \sigma^{(i-1)}$. We have

$$\Lambda' = \begin{vmatrix} \mu^{(i)} \\ \phi^{(i)} \\ \theta^{(i)} \\ \log(\sigma^{(i-1)}) \\ \log(c_u^{(i-1)}) \end{vmatrix} \sim N(\hat{\Lambda}, V)$$

From proposition 4, we have the conditional truncated distribution

$$\theta^{(i)}|\mu^{(i)}, \phi^{(i)}, c_u^{(i-1)}, \sigma^{(i-1)} \sim N(\hat{\Lambda}_3 + V_{3,-3}V_{-3,-3}^{-1}(\Lambda'_{-3} - \hat{\Lambda}_{-3}), \hat{\Lambda}_{3,3} - \hat{\Lambda}_{3,-3}\hat{\Lambda}_{-3,-3}^{-1}\hat{\Lambda}_{-3,3})$$

for $-1 \leq \theta^{(i)} \leq 1$.

ii. Check the feasibility of the draw [1]. If the solution is not feasible go to Step (i) and make another draw for $\theta^{(i)}$. Otherwise, find the likelihood $L_\theta^{(i)} = P(X_1^{(i-1)}, ..., X_T^{(i-1)}|\mu^{(i)}, \phi^{(i)}, \theta^{(i-1)}, c_u^{(i-1)}, \epsilon_0^{(i-1)}, X_0^{(i-1)})$. If $L_\theta^{(i)} > u_\theta$, then keep $\theta^{(i)}$ and repeat the same analysis for the next parameter($c_u^{(i)}$ in this case).

2. In this step, we generate the $X^{(i)}, \bar{X}^{(i)}|\Lambda^{(i)}, \Upsilon^{(i)}, \hat{X}$. This can be done in two ways:

(a) We can draw $X_t|\Lambda^{(i)}, X_{-t}^{(i)}$ by using a sequential slice sampler for $t = 1, ..., T$ where

$$X_{-t}^{(i)} = \left\{X_1^{(i)}, ..., X_{t-1}^{(i)}, X_{t+1}^{(i-1)}, ..., X_T^{(i-1)}\right\}$$

i. We compute the likelihood $L_X^{(i-1)} = P(\gamma^{(i-1)}; \upsilon^{(i-1)})$ where

$$\gamma_t(k)^{(i-1)} = \hat{X}_t(k)^{(i-1)} - \bar{X}_t(k)^{(i-1)}$$

[1] For K=1, the parameters are always feasible. For $K > 1$ we need the feasibility which means that $X^{(i-1)}$ and $\Lambda^{(i)}$ generates the optimal forecast $\bar{X}$

for $t = 1, ..., T$ and $k = 1, ..., K$.

This is our vertical slice.

ii. Draw a random variable u_X from Uniform$(0, L_X^{(i)})$. This is our horizontal slice.

iii. While true,

A. We can show that $X_t|\Lambda^{(i)}, X_{-t}^{(i)}$ follows a truncated normal distribution. Draw X_t from the truncated normal distribution.

B. Compute the likelihood $L_X^{(i)} = P(\gamma^{(i)}; \upsilon^{(i-1)})$. If $L_X^{(i)} > u_X$, then keep $X_t^{(i)}$ otherwise make another draw for X_t.

(b) This can also be done by drawing $\epsilon \sim N(0, \sigma^2 I)$ by using a slice sampler. In this case we draw ϵ together and compute the likelihood.

3. In this step, we generate $\Upsilon^{(i)}|\bar{X}^{(i)}, \hat{X}$. We first compute $\gamma^{(i)}$ as follows

$$\gamma_t(k)^{(i-1)} = \hat{X}_t(k)^{(i-1)} - \bar{X}_t(k)^{(i-1)}$$

for $t = 1, ..., T$ and $k = 1, ..., K$.

So we draw

- $p^{(i)}|\omega_1^{(i-1)}, \omega_2^{(i-1)}, \gamma^{(i)}$
- $\omega_1^{(i)}|p^{(i)}, \omega_2^{(i-1)}, \gamma^{(i)}$
- $\omega_2^{(i)}|p^{(i)}, \omega_1^{(i)}, \gamma^{(i)}$

by using a *sequential slice sampler*. Here, we show how we draw $p^{(i)}$. The same analysis is repeated for all the above parameters in the given order.

(a) We can compute the likelihood

$$L_p^{(i-1)} = P(\gamma^{(i)}|\Upsilon^{(i-1)})$$

This is our vertical slice.

(b) Draw a random variable u_p from Uniform$(0, L_p^{(i-1)})$. This is our horizontal slice.

(c) While true,

i. Draw $p^{(i)}$ from the prior $Beta(a, b)$.

ii. Find the likelihood $L_p^{(i)} = P(\gamma^{(i)}|p^{(i)}, \omega_1^{(i-1)}, \omega_2^{(i-1)})$. If $L_p^{(i-1)} > u_p$, then keep $p^{(i)}$ and repeat the same analysis for the next parameter($\omega_1^{(i)}$ in this case). Otherwise make another draw from the prior for $p^{(i)}$.

We also use shrinkage algorithm (Neal 2003) to improve the draws for the slice sampler. By doing that we decrease the number of rejected draws to find a feasible set of parameters.

4.3.3 Example

We use the customer forecasts in Example 10 to show the results of estimation. In Example 10, we only have forecasts for three weeks. By adding the forecasts for the following 17 weeks, we have the forecast vector $\hat{X}$ in Figure 4.4a. $\hat{X}$ provides $K = 5$ weeks forecasts in $T = 20$ weeks. We assume the following priors for estimation:

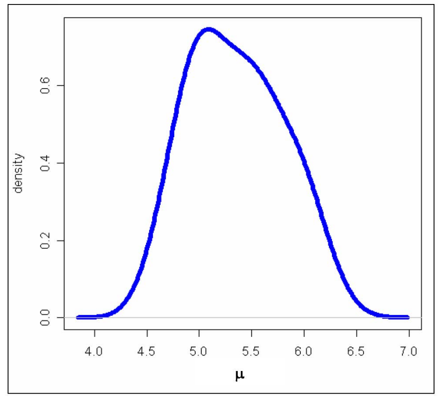
$$\begin{array}{lll} \mu \sim N(10,5) & \phi \sim N(0,0.5) & \theta \sim N(0,0.5) \\ \sigma \sim IG(2.5,10) & log(c_u) \sim N(0,1) & \\ \epsilon \sim N(0,\sigma) & X_0 \sim N(\mu, \sigma\sqrt{\frac{1+\theta^2-2\theta\phi}{1-\phi^2}}) & \end{array}$$

We run our sampler for 40,000 iterations to estimate the posterior distributions of each parameters. We discard the first 3,000 observations as the warmup period. The results are available in Figure 4.4. Some of the inferences that we can make from the posterior distributions are as follows:

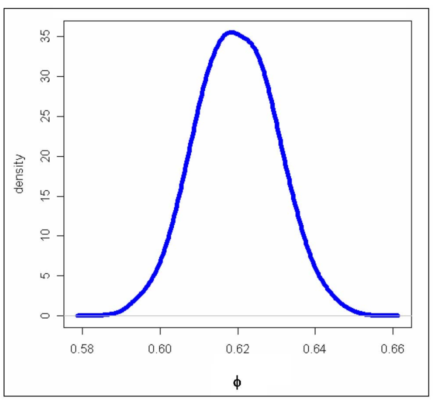
- We can see that the autoregressive parameter (ϕ) has a 95% confidence interval of $(0.59, 0.64)$. Therefore, there is a significant effect of the previous demand observation on current demand. However, for θ, the posterior centers around 0. We cannot reject the hypothesis that $\theta = 0$. This means that error terms are not autocorrelated. However, there is strong autocorrelation between the consecutive demand values.

$$\hat{X} = \left| \begin{array}{ccccc} 5 & 5 & 6 & 6 & 6 \\ 1 & 3 & 4 & 5 & 6 \\ 2 & 4 & 5 & 5 & 6 \\ 6 & 6 & 6 & 6 & 6 \\ 5 & 6 & 6 & 6 & 6 \\ 7 & 7 & 6 & 6 & 6 \\ 7 & 7 & 7 & 7 & 7 \\ 4 & 5 & 5 & 6 & 6 \\ 4 & 5 & 5 & 6 & 6 \\ 3 & 4 & 5 & 6 & 6 \\ 5 & 6 & 6 & 6 & 6 \\ 4 & 5 & 6 & 6 & 6 \\ 3 & 5 & 5 & 6 & 6 \\ 6 & 6 & 6 & 6 & 6 \\ 7 & 7 & 7 & 7 & 6 \\ 10 & 8 & 8 & 7 & 7 \\ 8 & 7 & 7 & 7 & 7 \\ 9 & 8 & 8 & 7 & 7 \\ 8 & 7 & 7 & 7 & 7 \\ 4 & 5 & 6 & 6 & 6 \end{array} \right|$$

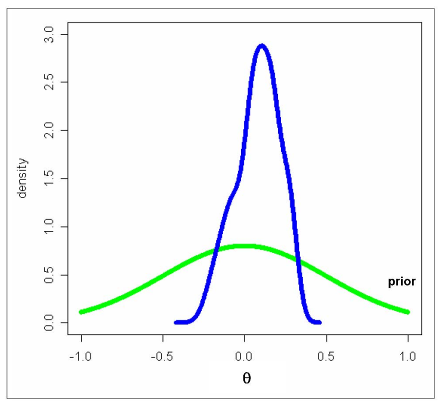
(a) $\hat{X}$



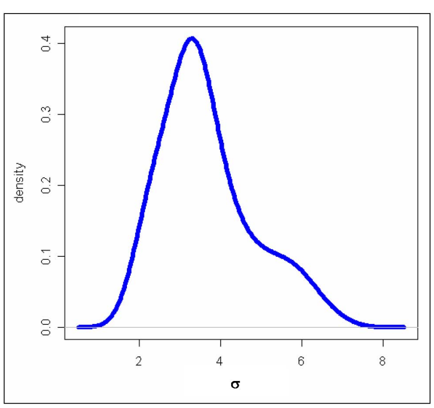
(b) 95% C.I.= $(4.61, 6.17)$



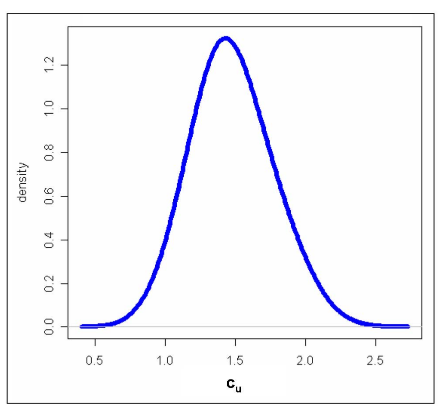
(c) 95% C.I.= $(0.59, 0.64)$



(d) 95% C.I.= $(-0.19, 0.29)$



(e) 95% C.I.= $(2.03, 6.31)$



(f) 95% C.I.= $(1.07, 1.96)$

Figure 4.4: The forecast vector $\hat{X}$ and the posterior distributions for μ, ϕ, θ, σ and c_u

- When we test the hypothesis that $c_u > c_o = 1$, we cannot reject it with $\alpha = 95\%$. This means that there is significant evidence from the data that the customer adds a positive bias to his forecasts.
- When we look at σ, we can see that there is high uncertainty at the customer. The high variance is an important factor that causes errors in the forecasts. However, σ by itself does not explain the forecast errors. As we discuss above, the customer adds a bias to forecasts due to his asymmetric loss function.

The supplier can estimate the expected unbiased forecast by looking at the data. For example at $t = 1$, the expected unbiased forecasts for the next five periods is available in Figure 4.5. Therefore, by looking at a history of 20 periods with 5-step ahead forecasts, the supplier can have significant information about the forecast generation model of the customer and can remove the bias from the forecast.

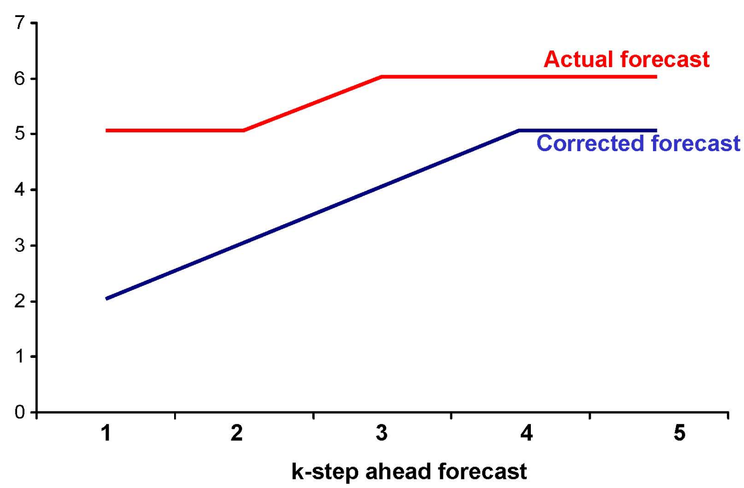


Figure 4.5: The actual and corrected forecast after the analysis at $t = 1$.

4.4 Hierarchical Model

In a hierarchical model, we study the diversity of the cost and demand parameters of the customers. Heterogeneity in cost and demand parameters give rise to different ordering behaviors from the customers. Depending on the diversity of the customers, hierarchical models provide better inferences about the forecast behavior of the customers. When the diversity is high, modeling the individual customer differences as well as the aggregate effects help the suppliers to better understand the forecast performance of the customers. In our analysis we have forecasts from multiple customers for multiple parts. Therefore, it is possible to make an inferences on the aggregate level as well as individual level. For example, we can analyze the forecasts of a customer for multiple parts and look at the forecast performance on an aggregate level. We can also perform an aggregate analysis for each part to understand the forecast performance of the manufacturers.

In our analysis we represent each customer with m and each part with n. We first define the first-stage priors as

$$\Lambda'_{mn} = \begin{vmatrix} \mu_{mn} \\ \phi_{mn} \\ \theta_{mn} \\ \log(\sigma)_{mn} \\ \log(c_u)_{mn} \end{vmatrix} \sim N(\hat{\Lambda}, V)$$

The second stage priors are

$$V \sim \text{Inverted Wishart}(\nu_0, S_0)$$

$$\hat{\Lambda} \sim N(0, \Lambda_0)$$

In order to have a full rank $\nu_0 > 5$ which is the number of parameters. Assume that we run an analysis over M different customers for a part. So we can drop n from the subscript in this case, we have

$$P(V) \propto |V|^{-\frac{\nu_0 - 6}{2}} etr\left(-\frac{1}{2} S_0 V^{-1}\right)$$

The posterior from the data is

$$\begin{aligned} P(V|\{\Lambda'_1, ..., \Lambda'_M\}) &\propto |V|^{-\frac{M+\nu_0-6}{2}} etr\left(-\frac{1}{2}(S_0 + S+)V^{-1}\right) \\ &= \text{Inverted Wishart}(\nu_0 + M, S_0 + S) \end{aligned}$$

where $S = \sum_{m=1}^{M} (\Lambda'_m - \bar{\Lambda}')(\Lambda'_m - \bar{\Lambda}')^T$.

The posterior for $\hat{\Lambda}$ from the data is

$$\begin{aligned} P(\hat{\Lambda}|V, \{\Lambda'_1, ..., \Lambda'_M\}) &\propto exp\left(\frac{1}{2}(\hat{\Lambda} - \hat{\Lambda}_M)^T V_M^{-1} (\hat{\Lambda} - \hat{\Lambda}_M)\right) \\ &= N(\hat{\Lambda}|\hat{\Lambda}_M, V_M) \end{aligned}$$

where

$$\begin{aligned} V_M &= \Lambda_0^{-1} + MV^{-1} \\ \hat{\Lambda}_M &= V_M^{-1} M V^{-1} \bar{\Lambda}' \end{aligned}$$

In order to predict the parameters, in iteration i we sequentially draw

1. For customers $m = 1, ..., M$, we draw $\Lambda_m^{'(i)}|\Lambda^{'(i-1)}$ by using the sampler in section 4.3.2.
2. We draw $V^{(i)}|\{\Lambda_1^{'(i)}, ..., \Lambda_M^{'(i)}\}$ from the posterior distribution $N(\hat{\Lambda}|\hat{\Lambda}_M^{(i)}, V_M^{(i)})$.
3. We draw $\hat{\Lambda}^{(i)}|V^{(i)}, \{\Lambda_1^{'(i)}, ..., \Lambda_M^{'(i)}\}$ from the posterior distribution Inverted Wishart$(\nu_0 + M, S_0 + S^{(i)})$.

4.4.1 Example

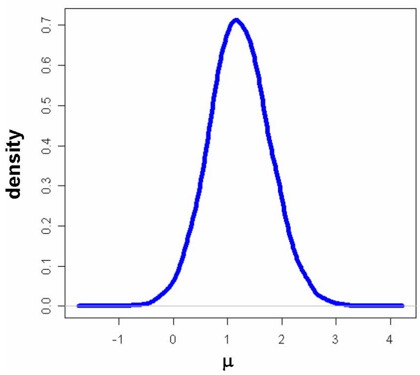
In Figure 4.6, we provide the estimates for the mean level $\hat{\Lambda}$ of the model parameters. We analyze a product with 16 customers. During our analysis, we observe that for most of the customers, our model represents a good fit and we do not observe too many errors in the

$$S_0 = \begin{bmatrix} 25 & 0 & 0 & 0 & 0 \\ 0 & 25 & 0 & 0 & 0 \\ 0 & 0 & 25 & 0 & 0 \\ 0 & 0 & 0 & 25 & 0 \\ 0 & 0 & 0 & 0 & 25 \end{bmatrix}$$

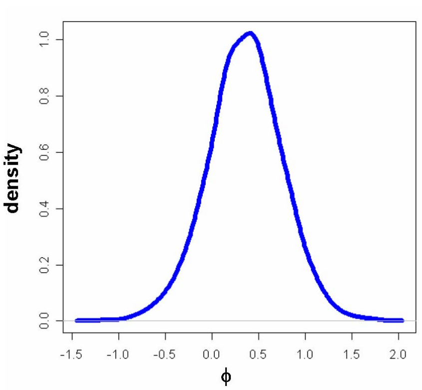
$$\nu_0 = 6$$

$$\Lambda_0 = \begin{bmatrix} 100 & 0 & 0 & 0 & 0 \\ 0 & 100 & 0 & 0 & 0 \\ 0 & 0 & 100 & 0 & 0 \\ 0 & 0 & 0 & 100 & 0 \\ 0 & 0 & 0 & 0 & 100 \end{bmatrix}$$

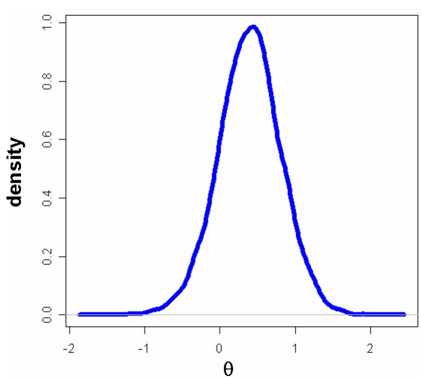
(a) The second stage priors



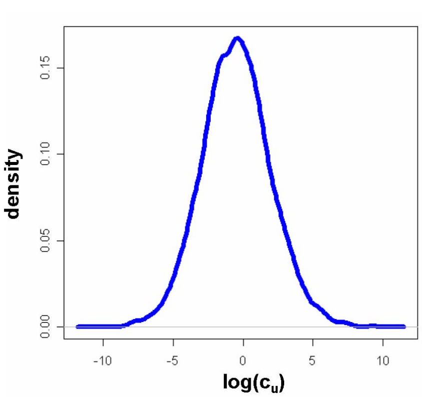
(b) Expected value for μ_i



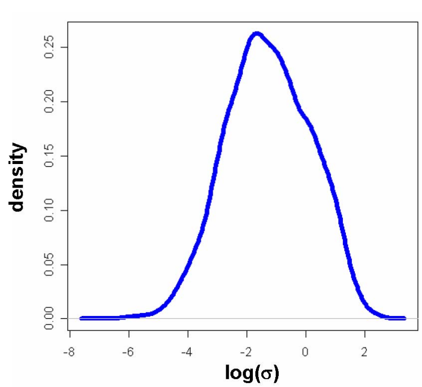
(c) Expected value for ϕ_i



(d) Expected value for θ_i



(e) Expected value for $\log(c_u)_i$



(f) Expected value for $\log(\sigma)_i$

Figure 4.6: The mean vector $\hat{\Lambda}$ for the hierarchical model with 16 customers and a single part

forecasts. For 6 customers, we observe that the forecasts are generated from a different process. Therefore, the estimates for the model parameters were not significant for those customers.

From 4.6, we have the following observations:

1. We can see that ϕ and θ tends to be greater than zero. This means that the autocorrelations between the forecasts and the errors of the consecutive periods tend to be positive.

2. The log value of cost of underestimation is very close to normal distribution with a mean value slightly less than zero. This means that underestimation is also a common behavior among the customers. This also shows that log transformation is necessary while analyzing the cost of overestimation and underestimation. At the individual level, we get more significant results for the customers where we cannot reject the hypothesis for customers over- or underestimating their orders.

3. The standard deviation of the demand tends to be small for some of the customers. For those customers, we can argue that the forecast errors are mostly due to the cost structure of the customer rather than the uncertainty in usage.

4.5 Conclusion and Future Research

We have used a time-series framework to model the evolution of forecast vectors. In our data, we observe that customers consistently overestimate or underestimate their orders. The goal of our analysis was to develop a framework to test the significance of strategic behavior. By using asymmetric loss functions, we explain the dynamics behind the bias in customer's forecasts. We present an estimation procedure which has high predictive power to understand the cost and demand structure of a customer from his forecasts.

Our hierarchical model can be extended in order to include individual parameters for each customer and part. We can also include exogenous parameters in order to test their effect on the forecast performance of the customers. The hierarchical models for multiple customers and parents are ideal for gibbs sampling methods where we sequentially draw individual parameters by conditioning on the other parameters. We mainly use the posterior distributions or slice sampling to draw the parameters. Other MCMC methods can also be considered for the analysis. The main advantage of the slice sampler is we do not need to tune the parameters and provide a proposal density.

We show that the customers add bias to their forecast due to their cost and demand structure. Therefore, it is critical for a supplier to create unbiased forecasts from the reported forecasts. Our analysis helps us understand which factors influence the forecast behavior of a manufacturer. A high σ for a customer shows that there is high uncertainty in the usage. A high c_u signals high cost of stockouts for the customer. We can also make inferences on the aggregate level. If we detect a common poor forecast performance across customers, there might be problems with shipping of the product which creates artificial spikes in the forecasts. Hierarchical model helps us understand how much of the errors can be explained from the individual and aggregate demand and cost structure.

The customer demand can follow non-ARMA processes. In this case we observe high ω_1 or ω_2 values, which show that the forecasts from the model do not fit the reported forecasts very well. Therefore, some other demand models can be considered for the non-ARMA customers. Our analysis can be extended to models with more components to capture these complexities. Possible such extensions are adding seasonality, production plans and nonlinear loss functions.

Chapter 5

Appendix

5.1 Appendix for Chapter 2

5.1.1 Sensitivity Analysis for the Single Period Problem

Part i: As stated in Proposition 2, there exists a S_{01} for a single period problem. When we increase Q to $Q + \triangle Q$ ($\triangle Q > 0$), the value of S_{01} may also change.

Case 1: $S_{01} \leq S_0$

When $S_{01} \leq S_0$, at point $x = S_{01}$,

$$L_1(x, x+Q) = L_0(x, S_0) \quad \Rightarrow \quad c_1 Q + H(x+Q) = c_0(S_0 - x) + H(S_0). \tag{5.1}$$

As ordering $Q + \triangle Q$ is dominated by ordering Q at $x = S_{01}$, we have

$$L_1(x, x+Q+\triangle Q) \quad > \quad L_1(x, x+Q). \tag{5.2}$$

Using Equations (5.1) and (5.2), at $x = S_{01}$, we get the following result:

$$L_1(S_{01}, S_{01}+Q+\triangle Q) > L_0(S_{01}, S_0). \tag{5.3}$$

At point $x = S_{01} - \triangle Q$, we have

$$L_0(S_{01} - \triangle Q, S_0) \quad = \quad c_0(S_0 - S_{01} + \triangle Q) + H(S_0) = c_0 \triangle Q + L_0(S_{01}, S_0), \tag{5.4}$$

$$L_1(S_{01} - \triangle Q, S_{01}+Q) \quad = \quad c_1(Q + \triangle Q) + H(S_{01}+Q) = c_1 \triangle Q + L_1(S_{01}, S_{01}+Q). \tag{5.5}$$

From Equation (5.1), we have $L_0(S_{01}, S_0) = L_1(S_{01}, S_{01}+Q)$. Since $c_1 \triangle Q < c_0 \triangle Q$, we have

$$\begin{aligned} c_1 \triangle Q + L_1(S_{01}, S_{01}+Q) \quad &< \quad c_0 \triangle Q + L_0(S_{01}, S_0) \\ L_1(S_{01} - \triangle Q, S_{01}+Q) \quad &< \quad L_0(S_{01} - \triangle Q, S_0). \end{aligned} \tag{5.6}$$

Therefore, from inequalities (5.3) and (5.6), we have the following results:

$$\begin{aligned} L_1(S_{01}, S_{01}+Q+\triangle Q) &> L_0(S_{01}, S_0), \\ L_1(S_{01}-\triangle Q, S_{01}+Q) &< L_0(S_{01}-\triangle Q, S_0). \end{aligned}$$

From the above inequalities, $L_1(x, x+Q+\triangle Q) < L_0(x, S_0)$ at $x = S_{01} - \triangle Q$ and $L_1(x, x+Q+\triangle Q) > L_0(x, S_0)$ at $x = S_{01}$. Since $L_1(x, x+Q+\triangle Q)$ and $L_0(x, S_0)$ are continuous functions, there should be another point $x = S_{01} + \triangle S_{01}$ in $S_{01} - \triangle Q < x < S_{01}$ where $L_1(x, x+Q+\triangle Q) = L_0(x, S_0)$. This point is the unique S_{01}^* for the new $Q^* = Q + \triangle Q$. Therefore, we have $S_{01} - \triangle Q < x = S_{01} + \triangle S_{01} < S_{01}$ which can be written as $-1 < \frac{\triangle S_{01}}{\triangle Q} < 0$.

Case 2: $S_{01} > S_0$

When $S_{01} > S_0$, we consider only the case with $S_0 < S_{01} < S_1$, since ordering Q cannot be optimal for $x \geq S_1$ for any Q values. At point $x = S_{01}$, we have $L_1(x, x+Q+\triangle Q) > L_1(x, x+Q)$. Since $L_1(S_{01}, S_{01}+Q) = L_0(S_{01}, S_{01})$, we get:

$$L_1(x, x+Q+\triangle Q) > L_0(x, x). \tag{5.7}$$

The cost of not ordering at $x = S_{01} - \triangle Q$ is $L_0(S_{01} - \triangle Q, S_{01} - \triangle Q) = H(S_{01} - \triangle Q)$. Since $x < S_1$, we have $\frac{\partial H(x)}{\partial x} < -c_1$. So, we have

$$\begin{aligned} H(S_{01}) - H(S_{01} - \triangle Q) &< -c_1 \triangle Q \\ H(S_{01}) + c_1 \triangle Q &< H(S_{01} - \triangle Q) \\ c_1 Q + H(S_{01}+Q) + c_1 \triangle Q &< H(S_{01} - \triangle Q) \\ L_1(S_{01} - \triangle Q, S_{01}+Q) &< L_0(S_{01} - \triangle Q, S_{01} - \triangle Q). \end{aligned} \tag{5.8}$$

The third inequality follows as $L_1(S_{01}, S_{01}+Q) = L_0(S_{01}, S_{01})$; i.e., $c_1 Q + H(S_{01}+Q) = H(S_{01})$. Therefore, at $x = S_{01} - \triangle Q$, we have $L_1(x, x+Q+\triangle Q) < L_0(x, x)$. We also have $L_1(x, x+Q+\triangle Q) > L_0(x,x)$ at $x = S_{01}$ from Equation (5.7). Therefore, there should be another point $x = S_{01} + \triangle S_{01}$ which satisfies $S_{01} - \triangle Q < x < S_{01}$ and $L_1(x, x+Q+\triangle Q) = L_0(x,x)$. This point is the unique S_{01}^* for the new $Q^* = Q + \triangle Q$. So we have $-1 < \frac{\triangle S_{01}}{\triangle Q} < 0$.

Part ii: When $S_{01} \leq S_0$, at point $x = S_{01}$, $L_1(x, x+Q) = L_0(x, S_0)$. When we increase the discounted price c_1 to $c_1^* = c_1 + \triangle c_1$ ($\triangle c_1 > 0$), let the new critical level be $S_{01}^* = S_{01} + \triangle S_{01}$ and $L_{1^*}(x,y)$ be the cost of ordering with price c_1^*. At point $x = S_{01}$, we have $L_{1^*}(x, x+Q) > L_1(x, x+Q) = L_0(x, S_0)$. Therefore, with the new price c_1^*, ordering with the discount is no longer optimal at point $x = S_{01}$. Since the order quantities are monotonic in initial inventory position (Proposition 2), there should be another $S_{01}^* < S_{01}$. Therefore, $\frac{\partial S_{01}}{\partial c_1} < 0$. The case $S_{01} > S_0$ follows from the above discussion by changing $L_0(x, S_0)$ with $L_0(x,x)$.

Part iii: When $S_{01} \geq S_0$, we do not observe S_0. Therefore, changing c_0 does not have any effect on S_{01} and $\frac{\partial S_{01}}{\partial c_0} = 0$. Next, we consider the case $S_0 > S_{01}$. When we increase the original price c_0 to $c_0^* = c_0 + \triangle c_0$ ($\triangle c_0 > 0$), the values of S_0 and S_{01} may also change. Let the new values of S_{01} and S_0 be $S_{01}^* = S_{01} + \triangle S_{01}$ and $S_0^* = S_0 + \triangle S_0$, respectively. Let $L_{0^*}(x,y)$ be the cost of ordering with price c_0^*. Since ordering the same quantities with c_0 is less costly, we have $L_0(x, S_0) \leq L_0(x, S_0^*) < L_{0^*}(x, S_0^*)$. At point $x = S_{01}$, we have $L_1(x, x+Q) = L_0(x, S_0) < L_{0^*}(x, S_0^*)$. Therefore, ordering with the discount is a better alternative at $x = S_{01}$. Since the order quantities are monotonic in initial inventory position (Proposition 2), there should be another S_{01}^* which satisfies $S_{01}^* > S_{01}$, i.e., $\triangle S_{01} > 0$. Therefore, $\frac{\partial S_{01}}{\partial c_0} > 0$.

5.1.2 Analysis of Order Quantities in an Ordering Interval

For any x in an ordering interval s, the buyer has two options. He can either order $q \geq Q$ and get the discount or order up to a level without getting the discount. There may be many optimal order-up-to levels in a given interval, but there will be at most one S_{01}^{s}. We prove this by contradiction.

We show that there does not exist any inventory position x_c in an ordering interval such that, in the optimal policy, the buyer orders with the discount when $x > x_c$ and orders with the original price when $x \leq x_c$. The proof involves three steps:

1. The buyer's order-up-to levels with and without the discount do not change with infinitesimal increases in the inventory position if the original order-up-to level is still feasible.

2. If the buyer orders without the discount when the inventory position is x, he will not order $q > Q$ with the discount when the inventory position is $x + \Delta$, where $\Delta > 0$.

3. If the buyer orders without the discount when the inventory position is x, he will not order $q = Q$ with the discount when the inventory position is $x + \Delta$, where $\Delta > 0$.

For base-stock levels to exist[1], $H_n(x)$ function should be continuous, which is the case in our problem for $n = 1, ..., N$. For price c_i, let the order-up-to level for x be S_i^1. When the optimal order quantity $q^*(x) > 1_{\{c_i=c_1\}}Q$, where $1_{\{c_i=c_1\}}$ is the indicator function[2], at $x + \Delta$ for infinitesimal $\Delta > 0$ the buyer should order $q^*(x) - \Delta$ to order up to S_i^1. If the

[1] We prove our theorem for finite horizon with any number of periods. Since the optimal policy for infinite horizon is derived by using policy iteration, the same result holds for infinite horizon.

[2] This condition guarantees that for an infinitesimal Δ, the optimal order-up-to level for x is still feasible (but not necessarily optimal) for $x + \Delta$.

order-up-to level for $x + \Delta$ is S_i^2, s.t. $S_i^2 \neq S_i^1$, then

$$\begin{aligned}
L_i(x+\Delta, S_i^1) &> L_i(x+\Delta, S_i^2) \\
c_i(S_i^1 - x - \Delta) + H_n(S_i^1) &> c_i(S_i^2 - x - \Delta) + H_n(S_i^2) \\
c_i(S_i^1 - x) + H_n(S_i^1) &> c_i(S_i^2 - x) + H_n(S_i^2) \\
L_i(x, S_i^1) &> L_i(x, S_i^2). \qquad (5.9)
\end{aligned}$$

So, ordering up to S_i^2 is optimal for x, which is a contradiction. Therefore, if $q^*(x) > 1_{\{c_i=c_1\}}Q$, the order-up-to level does not change with an infinitesimal increase in the inventory level.

Next we prove Steps 2 and 3 by considering two different cases. In these steps, we assume that for the original price, x and $x + \Delta$ are in the same ordering interval; i.e., x is strictly less than the order-up-to level of the original price at that point.

Case 1: Ordering more than Q to get the discount

In this first case, when the inventory position is x, in the optimal policy, the buyer orders up to S_0^1 without the discount and when the inventory position is $x + \Delta$, he orders up to S_1^1 and gets the discount. That being optimal, we have the following inequalities:

$$\begin{array}{llllll}
 & & \text{order without the discount} & & \text{order with the discount} \\
\text{at point } x & \Longrightarrow & c_0(S_0^1 - x) + H_n(S_0^1) & \leq & c_1(S_1^1 - x) + H_n(S_1^1) \\
\text{at point } x+\Delta & \Longrightarrow & c_0(S_0^1 - x - \Delta) + H_n(S_0^1) & \geq & c_1(S_1^1 - x - \Delta) + H_n(S_1^1).
\end{array}$$

Notice that S_1^1 is still a feasible order-up-to level for x, however, it may not be the best order-up-to level. If we multiply the second inequality by (-1), we get the following inequalities:

$$\begin{aligned}
c_0(S_0^1 - x) + H_n(S_0^1) &\leq c_1(S_1^1 - x) + H_n(S_1^1) \\
-c_0(S_0^1 - x - \Delta) - H_n(S_0^1) &\leq -c_1(S_1^1 - x - \Delta) - H_n(S_1^1).
\end{aligned}$$

When we sum up these two inequalities, we get $c_0 \leq c_1$, which contradicts with our cost assumptions. Therefore, if the buyer is not using the discount for an inventory position, he would never order a quantity greater than Q as the inventory position increases.

Case 2: Ordering Q to get the discount

In this second case, in the optimal policy, when the inventory position is x, the buyer orders up to S_0^1 without the discount and when the inventory position is $x + \Delta$, he orders Q and gets the discount. That being optimal, we have the following inequalities:

$$\begin{array}{llccc} & & \text{order without the discount} & & \text{order with the discount} \\ \text{at point } x & \Longrightarrow & c_0(S_0^1 - x) + H_n(S_0^1) & \leq & c_1 Q + H_n(x+Q) \\ \text{at point } x+\Delta & \Longrightarrow & c_0(S_0^1 - x - \Delta) + H_n(S_0^1) & \geq & c_1 Q + H_n(x+\Delta+Q) \end{array}$$

Multiply the second inequality by (-1),

$$c_0(S_0^1 - x) + H_n(S_0^1) \;\; \leq \;\; c_1 Q + H_n(x+Q) \tag{5.10}$$

$$-c_0(S_0^1 - x - \Delta) - H_n(S_0^{n'}) \;\; \leq \;\; -c_1 Q - H_n(x+\Delta+Q) \tag{5.11}$$

and sum up these two inequalities:

$$c_0 \leq \frac{H_n(x+Q) - H_n(x+\Delta+Q)}{\Delta}. \tag{5.12}$$

As $c_1 < c_0$,

$$\begin{aligned} c_1 &< \frac{H_n(x+Q) - H_n(x+\Delta+Q)}{\Delta} \\ c_1\Delta &< H_n(x+Q) - H_n(x+\Delta+Q) \\ c_1\Delta + H_n(x+\Delta+Q) &< H_n(x+Q). \end{aligned}$$

Adding $c_1 Q$ to both sides, we get

$$c_1(Q + \Delta) + H_n(x + \Delta + Q) \leq c_1 Q + H_n(x + Q). \tag{5.13}$$

Therefore, at x ordering up to a level $x + \Delta + Q$ is better than ordering Q and we can get a tighter bound on inequality (5.10):

$$c_0(S_0^1 - x) + H_n(S_0^1) \quad \leq \quad c_1(Q + \Delta) + H_n(x + \Delta + Q).$$

We can use this inequality with inequality (5.11) to reach at the following result:

$$\begin{array}{llcl} \text{at point } x & \Longrightarrow \quad c_0(S_0^1 - x) + H_n(S_0^1) & \leq & c_1(Q + \Delta) + H_n(x + \Delta + Q) \\ \text{at point } x + \Delta & \Longrightarrow \quad -c_0(S_0^1 - x - \Delta) - H_n(S_0^1) & \leq & -c_1 Q - H_n(x + \Delta + Q) \end{array}$$

When we sum up these two inequalities, we get $c_0 \leq c_1$, which contradicts with our cost assumptions. Therefore, if the buyer is not using the discount for an inventory position, he would never order a quantity equal to Q as the inventory position increases. The buyer would never switch from not using the discount to using the discount in an ordering interval. Therefore, there is at most one critical level S_{01}^s in an ordering interval s.

5.1.3 N-Period Problem

In this section, we study the structural properties of the finite horizon planning problem. The buyer's objective is to minimize his long-run average total cost. In this problem, the terminal costs should be taken into account. As we have already discussed, the terminal penalty cost $\bar{p}$ is larger than the original price; i.e., $\bar{p} > c_0$. To incorporate the salvage cost, we update holding cost parameter h of the last period without losing the convexity of the last period's cost function. We assume that the terminal holding cost $\bar{h}$, i.e., salvage cost, is more than $-h$, which gives a positive updated holding cost. In our computational analysis,

p and h are taken to be constant[3].

Figure 5.1 displays the first-period optimal policy for a two-period problem. Note that, in the first period, a three-index policy is *not* optimal and order quantities are *not* monotonic in the initial inventory position. The buyer changes his procurement strategy based on his initial inventory position. If the buyer is in a backorder situation at the beginning of the first period (region A), then he prefers to procure only for the first period using the quantity discount. When he is in region B, the buyer inflates his orders in order to get the discount by ordering exactly Q. If the buyer has very little inventory at the beginning of first period (region C), he procures for both periods in order to take advantage of the discount opportunity. Region C destroys the monotonicity of the policy. This behavior is similar to *buy-and-hold* strategy that is introduced in Section 2.4.1. As his inventory level increases further, the buyer either procures without quantity discount or he does not order at all (region D). To summarize, for the finite horizon problem the optimal policy can be quite complicated.

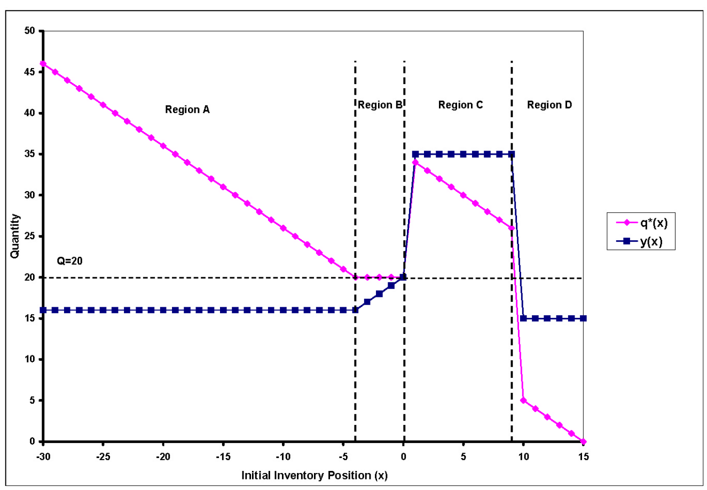


Figure 5.1: For Discrete Uniform[16, 20], with $c_0 = 1,\ c_1 = 0.7,\ p = 0.4,\ h = 0.15,\ \bar{p} = 2,\ \bar{h} = 0.3$, and $Q = 20$, the first-period optimal policy for the two-period problem.

[3]For N-period problem, we can relax the stationarity assumption as well as constant p and h values. The same analysis would extend to identical distributions, and nonidentical p and h values.

5.1.4 Effect of Increasing Q on Minimum Order Quantity Problem

Let $P_1(x)$ be the optimal policy, $P_2(x)$ be the best two-index policy, and $P_3(x)$ be the best one-index policy, where the buyer orders exactly Q units each time his initial inventory position falls below a critical level $\hat{S}_{01}$. We use these three policies to prove Theorem 3 in four steps: (1) In the optimal policy $P_1(x)$, for $x > \hat{S}_{01}$, the buyer does not place an order. (2) The buyer's consecutive orders become less dependent on each other as Q increases. (3) The order quantities of $P_2(x)$ converges to $P_3(x)$ as Q goes to infinity. (4) Two-index policy $P_2(x)$ becomes optimal when Q is large enough. Therefore, $P_3(x)$ is optimal in the limit.

Step 1: For policy $P_3(x)$, the critical level $\hat{S}_{01}$ is the inventory position where the buyer is indifferent between ordering Q and waiting till the next period to order Q. The expected inventory position after the next period is the same for both actions. The only difference is in the expected penalty and holding costs for the current period which have the functional form $H(x)$ of a single period.

- $H(x+Q) - H(x)$ is increasing in x.
- Since $H(x)$ is a convex function, there is a unique critical level $x = \hat{S}_{01}$ where $H(x+Q) = H(x)$. As Q goes to ∞, $x = \hat{S}_{01}$ goes to $-\infty$[4].
- From Proposition 3, we know that $\frac{\partial \hat{S}_{01}}{\partial Q} > -1$, i.e., $\hat{S}_{01} + Q$ is increasing with Q. For a large enough Q, $Pr(u > \hat{S}_{01} + Q) \approx 0$ where u is the demand for the next period. Therefore, $H(\hat{S}_{01} + Q) \approx (Q + \hat{S}_{01} - \mu)h$.

Using the two observations above for large Q values, we get the following:

$$H(\hat{S}_{01} + Q) - H(\hat{S}_{01}) \; = \; 0 \Rightarrow (Q + \hat{S}_{01} - \mu)h \;\approx\; (-\hat{S}_{01} + \mu)p \Rightarrow \hat{S}_{01} \;\approx\; \mu - Q\left(\frac{h}{h+p}\right),$$

which is an equality in the limit as $Pr(u > \hat{S}_{01} + Q)$ goes to zero. Therefore, the critical level $\hat{S}_{01}$ for $P_3(x)$ is $\hat{S}_{01} = \mu - Q\left(\frac{h}{h+p}\right)$.

[4] From the sensitivity analysis in Proposition 3, $\hat{S}_{01}$ satisfies $\frac{\partial \hat{S}_{01}}{\partial Q} < 0$. We assume a very large c_0 value, such that the buyer does not order with the original price. As $-p \leq \frac{dH(x)}{dx} \leq h$, increasing Q decreases $\hat{S}_{01}$ down to $-\infty$ in the limit.

One can show that in the optimal policy $P_1(x)$, for any ordering interval, $S_{01} \leq \hat{S}_{01}$ and $S_{01} \rightarrow \hat{S}_{01}$ as $Q \rightarrow \infty$. Assume that this is not the case. Then for $\hat{S}_{01} \leq x \leq S_{01}$, in the optimal policy, there exists a region where the buyer orders Q. Compare two solutions: ordering Q now and ordering nothing in this period, but Q in the next period. The latter is a feasible solution, but may not necessarily be optimal for the buyer. Using a similar analysis to the one above, we can show that the former action is dominated by the latter. Hence, in the optimal policy the buyer postpones his order, which is a contradiction. Therefore, $S_{01} \leq \hat{S}_{01}$.

Step 2: From step 1 we know that in the optimal policy $P_1(x)$ the buyer can order only when $x \leq \hat{S}_{01}$. We also show that $\hat{S}_{01}$ is a decreasing linear function of Q when Q is very large. When the buyer orders in $P_1(x)$, he always orders up to an inventory position larger than $\tilde{S}_1$, which is the minimum of the order-up-to levels. We can bound $\tilde{S}_1$ from below by the infinite horizon solution for $Q = 0$[5]. Therefore, as Q increases, the region where the buyer does not place an order (includes x: $\hat{S}_{01} < x \leq S_1^{Q=0} < \tilde{S}_1$) and the time till the next order get very large (in the limit the time till the next order goes to ∞), and the orders become independent from each other for $P_1(x)$.

Step 3: In this step, we will prove that as Q goes to infinity, the region where the buyer orders more than Q disappears; i.e., $P_3(x) = P_2(x)$[6]. Let $\hat{S}_1$ be the order-up-to level which is not exercised by policy $P_3(x)$. In step 1, we show that the buyer does not order at all when $x > \hat{S}_{01}$. When $x \leq \hat{S}_1 - Q$, the buyer orders more than Q in the optimal policy, but not in $P_3(x)$. Therefore, $P_3(x)$ is not optimal for finite Q. However, as Q goes to infinity, $\hat{S}_1 - Q$ goes to $-\infty$. In the limit, the region $x < \hat{S}_1 - Q$ where the buyer orders more than

[5]The order-up-to levels are higher for $Q > 0$ compared to $Q = 0$. To see why, assume the opposite. Let $S_1^{Q=0}$ be the order-up-to level for $Q = 0$ and $S_1^{Q>0}$ be the order-up-to level for $Q > 0$. For $S_1^{Q>0} - Q \leq x \leq S_1^{Q=0} - Q$ ordering up to $S_1^{Q=0}$ is feasible for $Q > 0$ and it is a better alternative. Therefore, $S_1^{Q>0}$ cannot be the optimal order-up-to level.

[6]All the following expressions are satisfied as an equality only when Q is large enough.

Q disappears and $P_3(x)$ is optimal.

Assume that the buyer follows $P_3(x)$. The buyer orders up to $\hat{S}_{01} + Q = \mu + Q\left(\frac{p}{h+p}\right)$ at $x = \hat{S}_{01}$. For any initial inventory $x \leq \hat{S}_{01}$, the inventory position after the buyer places an order is $x + Q$. The maximum value of the average inventory that the buyer holds is when $x = \hat{S}_{01}$ and is equal to $Q\left(\frac{p}{2(h+p)}\right)$ as the expected inventory position after the first period of the cycle is $Q\left(\frac{p}{h+p}\right)$. Therefore, we can represent the average inventory when the buyer places any order at x as a function of this maximum value; it can be written as $wQ\left(\frac{p}{2(h+p)}\right)$ where $0 < w \leq 1$. The average number of items backlogged is independent of the previous order size as discussed in step 2 and equal to $Q\left(\frac{h}{2(h+p)}\right)$ $\left(\text{as } \hat{S}_{01} = \mu - Q\left(\frac{h}{(h+p)}\right), \text{ the maximum backlog quantity is } Q\left(\frac{h}{2(h+p)}\right)\right)$.

When the buyer orders Q at initial inventory position x, this starts a new order cycle. We call the expected number of periods in this order cycle with on-hand inventory and backlog $t^+(x+Q)$ and $t^-(x+Q)$, respectively. The expected number of periods in an order cycle is $t(x+Q) = t^+(x+Q) + t^-(x+Q)$. For Q large enough, $\frac{t^+(x+Q)}{t^-(x+Q)} = \frac{wQ\frac{p}{\mu(h+p)}}{Q\frac{h}{\mu(h+p)}} = \frac{wp}{h}$, where μ is the mean demand. Therefore, the expected penalty cost in an order cycle is the multiplication of the average number of periods with backorders $t^-(x+Q)$, penalty cost p, and the average backorder $Q\left(\frac{h}{2(h+p)}\right)$. Similarly, the expected holding cost in an order cycle is the multiplication of the average number of periods with inventory $t^+(x+Q)$, holding cost h, and the average on-hand inventory $wQ\left(\frac{p}{2(h+p)}\right)$. Therefore, an approximation for the expected average cost per period in an order cycle that starts with an initial inventory position x is as follows:

$$\begin{aligned} C(w) &= \frac{p \cdot t^-(x+Q) \cdot Q\left(\frac{h}{2(h+p)}\right) + h \cdot t^+(x+Q) \cdot wQ\left(\frac{p}{2(h+p)}\right)}{t^-(x+Q) + t^+(x+Q)} \\ &= Q\frac{ph}{2(h+p)}\frac{t^-(x+Q) + wt^+(x+Q)}{t^-(x+Q) + t^+(x+Q)} \\ &= Q\frac{ph}{2(h+p)}\frac{t^-(x+Q)\left(1+\frac{pw^2}{h}\right)}{t^-(x+Q)(1+\frac{pw}{h})} \quad = \quad Q\frac{hp}{2(h+p)}\left(\frac{h+pw^2}{h+pw}\right). \end{aligned}$$

The average cost[7] $C(w)$ has a minimum at $w^* = -\frac{h}{p} + \sqrt{\left(\frac{h}{p}\right)^2 + \frac{h}{p}} < 1$ which satisfies $0 < w^* < 1$.

For $w > w^*$, the average cost decreases as w decreases. Therefore, the buyer prefers to order the minimum quantity possible, which is Q. When $w \leq w^*$, the average cost decreases as w increases. Therefore, the buyer prefers to order more than Q. Therefore, when $w = w^*$, which corresponds to $x = w^*Q\left(\frac{p}{h+p}\right) - Q + \mu = \hat{S}_1 - Q$, the buyer would do better than $P_3(x)$. Hence, the interval where the buyer orders exactly Q units is $\left(\hat{S}_1 - Q, \hat{S}_{01}\right)$ and the length of this interval is

$$\begin{aligned} \hat{S}_{01} - (\hat{S}_1 - Q) &= \hat{S}_{01} - w^*Q\frac{p}{(h+p)} + Q - \mu \\ &= \mu - Q\frac{h}{(h+p)} - w^*Q\frac{p}{(h+p)} + Q - \mu \;=\; Q\frac{h}{(h+p)}(1-w^*) \;=\; Q \cdot \varphi(h,p) \end{aligned}$$

where $\varphi(h,p)$ is a positive function of p and h. The length of the interval, $\hat{S}_{01} - (\hat{S}_1 - Q)$, is a linear increasing function of Q. Hence, Q can be made arbitrarily large in order to make the probability of placing an order larger than Q, $Pr(u > S_{01} - (S_1 - Q)) = \alpha$, close to zero; i.e., as Q increases, the probability of getting to an inventory position where $P_3(x)$ is not optimal goes to zero. Therefore, as Q goes to infinity, α becomes zero and $P_2(x)$ converges to $P_3(x)$.

Step 4: Assume that at period $k : 0 < k < \infty$ two-index policy $P_2(x)$ is optimal. If we can show that $P_2(x)$ stays optimal for period $k+1$, then we prove that two-index policy is the optimal policy of the infinite horizon problem by using policy iteration (Bertsekas 1995). From step 1, we know that $S_{01} \leq \hat{S}_{01}$. From step 3, we know that as Q goes to infinity, $P_2(x)$, S_1, and S_{01} converge to $P_3(x)$, $\hat{S}_1$, and $\hat{S}_{01}$ respectively. Using the result of step 2, the decision in period $k+1$ is independent from the next order placed and only

[7] The buyer orders up to $\hat{S}_{01} + Q = \mu + Q\left(\frac{p}{h+p}\right)$ at $x = \hat{S}_{01}$. As Q goes to infinity, $Q\left(\frac{p}{h+p}\right)$ gets really large and w converges to 1. Therefore, the average cost becomes $C(1) = \frac{Qph}{2(h+p)}$ in the limit.

depends on the current order to be placed. Therefore, the same policy in k can be shown to be optimal for $k+1$. A two-index policy is irreducible which is a necessary condition for having a stationary policy. Hence, $P_2(x)$ is optimal for the infinite horizon problem for Q large enough.

5.1.5 Effect of Increasing Discount Rate on Quantity Discount Problem

In our analysis, we fix the price c_1. We adjust c_0 in order to provide a discount ratio. A policy $P()$ defines the optimal order quantities for each initial inventory position. The buyer exercises all the prices stated in the policy for at least one initial inventory position. For example, $P(c_1, c_0)$ is the optimal policy which exercises both prices c_0 and c_1 and $P(c_1)$ is the optimal policy which only exercises price c_1. The cost of the policy $P(c_1, c_0)$ is $L(P(c_1, c_0))$. Notice that, even though p, h, and Q do not appear in the notation, these parameters affect the optimal policy.

We first show that there exits a price $c_0 = c_0^1$, where $P(c_1, c_0)$ cannot be an optimal policy; i.e., $P(c_1)$ is the optimal policy for $c_0 = c_0^1$. The policy $P(c_1)$ has a finite cost $L(P(c_1))$. For all inventory positions where ordering with c_0 is optimal, we increase the cost c_0 large enough to prevent the buyer from ordering with price c_0 for that initial inventory position. Since $P(c_1)$ is a feasible solution, the buyer switches to that policy for all initial inventory positions for this large enough $c_0 = c_0^1$ value. Therefore, there always exists $c_0 = c_0^1$, where $P(c_1, c_0^1)$ cannot be an optimal policy.

Next, we show the existence of c_0^* which defines z^*. Assume that, starting from $c_0 = c_1$, we increase c_0 value. $P(c_0)$ is the optimal policy at $c_0 = c_1$. As we increase c_0, the buyer can exercise both prices c_0 and c_1 with policy $P(c_0, c_1)$. Once, c_0 hits c_0^*, the buyer exercises only price c_1 with optimal policy $P(c_1)$. So, we have

$$L(P(c_1)) \leq L(P(c_1, c_0^*)). \tag{5.14}$$

Now we prove that there does not exist any price $c_0 = c_0' > c_0^*$, with $P(c_1, c_0')$ as the optimal policy. Assume that $P(c_1, c_0')$ is optimal. So,

$$L(P(c_1, c_0')) \leq L(P(c_1)). \tag{5.15}$$

$P(c_1, c_0')$ is a feasible policy for price $c_0 = c_0^*$. The cost of using policy $P(c_1, c_0')$ when $c_0 = c_0^*$ is $L_{c_0^*}(P(c_1, c_0'))$. Since the buyer pays a lower price each time he orders with the original price, we have

$$L_{c_0^*}(P(c_1, c_0')) < L(P(c_1, c_0')). \tag{5.16}$$

Since $P(c_1)$ is optimal policy for $c_0 = c_0^*$, we also have

$$L(P(c_1)) \leq L_{c_0^*}(P(c_1, c_0')). \tag{5.17}$$

Combining inequalities (5.15), (5.16), and (5.17), we have

$$L(P(c_1)) < L(P(c_1)), \tag{5.18}$$

which is a contradiction. Therefore, for $c_0 > c_0^*$, $P(c_1)$ is the optimal policy.

At the beginning of the proof, we have shown that there always exist c_0^1 where $P(c_1)$ is optimal. Since $c_0^* \leq c_0^1$, there always exist a c_0^*. Therefore, for a given c_1 value, we can always attain a discount ratio z^* stated in the theorem.

5.1.6 Stochastic Quantity Discounts with the Disposal Option

In this section, we study quantity discounts with inventory disposal at the beginning of each period. The buyer may dispose his inventory for a unit cost of s_d, $s_d \geq 0$, after he places an order and before he satisfies the demand and incurs the inventory holding cost. We keep all the other modelling assumptions the same as our original model.

Single Period Problem with Inventory Disposal

We first define the single period problem and present the expected cost minimizing solution. The buyer chooses an order quantity and a disposal quantity before realizing demand. Let the starting inventory position be x. If the buyer orders up to y_1 with a unit price of c_j, $j = 0, 1$, and disposes inventory down to y_2 with a unit cost of s, the total cost $L_j(x, y_1, y_2)$ is

$$L_j(x, y_1, y_2) = c_j(y_1 - x) + s_d(y_1 - y_2) + H(y_2), \tag{5.19}$$

such that $y_1 \geq x$ and $y_1 \geq y_2$. Furthermore, if the buyer orders with the discounted price, his order quantity should be greater than Q, i.e., $y_1 \geq x + Q$. $H(y_2)$ is the penalty and holding cost function,

$$H(y_2) = \bar{p} \int_{y_2}^{\infty} (u - y_2) dF(u) + \bar{h} \int_{0}^{y_2} (y_2 - u) dF(u), \tag{5.20}$$

and it is convex in y_2. Therefore, $L_j(x, y_1, y_2)$ is convex in y_2. The theorem below presents the optimal ordering policy for the single period problem with inventory disposal.

Theorem 5 *The optimal order and disposal quantities $(q^*(x), d^*(x))$ for the single period all-unit quantity discount problem with one price break and inventory disposal is given by the following rule:*

$$(q^*(x), d^*(x)) = \begin{cases} (\max\{S_1 - x, Q\}, \max\{x + Q - D, 0\}) & \text{when} \quad x \leq S_{01}^d \\ (\max\{S_0 - x, 0\}, 0) & \text{when} \quad S_{01}^d < x \leq D \\ (0, x - D) & \text{when} \quad D < x \end{cases}$$

where $S_i = F^{-1}\left(\frac{\bar{p} - c_i}{\bar{h} + \bar{p}}\right)$. We call this policy a **four-index** *policy with indices $\left(S_0, S_1, S_{01}^d, D\right)$.*

Proof Sketch: Since $L_j(x, y_1, y_2)$ is convex in y_2, there is a unique D which we can find by taking the derivative of $L_j(x, y_1, y_2)$ with respect to y_2:

$$D = F^{-1}\left(\frac{\bar{p} + s_d}{\bar{h} + \bar{p}}\right). \tag{5.21}$$

The buyer disposes inventory down to D when $y_1 > D$. Then, there are two cases to analyze: (1) $y_1 > D$ and the buyer disposes inventory and (2) $y_1 \leq D$ and the buyer does not dispose inventory. In the former case the cost function is $L_j(x, y_1, D)$ and in the latter case the cost function is $L_j(x, y_1, y_1)$.

As $L_j(x, y_1, y_1)$ is a convex function of y_1, we can find the unique S_0 and S_1 that minimize the functions $L_0(x, y_1, 0)$ and $L_1(x, y_1, 0)$, respectively,

$$S_j = F^{-1}\left(\frac{\bar{p} - c_j}{\bar{h} + \bar{p}}\right) \qquad j = 0, 1. \tag{5.22}$$

We have $S_0 \leq S_1 \leq D$. Therefore, the buyer's order up to levels satisfy the feasibility condition of this case, i.e., $y_1 \leq D$. Hence, when the buyer orders up to S_0 or S_1, he does not dispose inventory. He disposes inventory only if his initial inventory position is too high and he does not order at all, i.e., $x > D$, and *may* dispose inventory when he orders exactly Q units.

For disposal to be an option in the latter case, $S_{01} + Q > D$, where S_{01} is the critical level of a three-index policy for the problem without disposal. When $S_{01} + Q > D$, the buyer's cost function can be written as

$$L_1(x, x + Q, x + Q - D) = c_1 Q + s_d(x + Q - D) + H(D), \tag{5.23}$$

which is linear in x. We can show that there exists a $S_{01}^d \geq S_{01}$ where the buyer stops ordering Q units and then disposing down to D. If $S_{01} + Q \leq D$, $S_{01}^d = S_{01}$. $\square$

We display the structure of the optimal policy for the single period problem with an example.

Example 11 *For Poisson*(6)*, with* $s_d = 0$*,* $c_0 = 1.0$*,* $c_1 = 0.5$*,* $\bar{p} = 1.5$*,* $p = 0.35$*,* $\bar{h} = 0.4$*, and* $Q = 20$*, the optimal policy for the single problem is as follows*[8]*:*

		initial inventory	$\leq$	*-14*	:	*order up to* **6***, dispose 0*
-13	$\leq$	*initial inventory*	$\leq$	*-12*	:	*order* $Q = 20$*, dispose 0*
-11	$\leq$	*initial inventory*	$\leq$	**-4**	:	*order* $Q = 20$*, dispose down to* **8**
-3	$\leq$	*initial inventory*	$\leq$	*4*	:	*order up to* **4***, dispose 0*
5	$\leq$	*initial inventory*	$\leq$	*8*	:	*order 0, dispose 0*
9	$\leq$	*initial inventory*			:	*order 0, dispose down to* **8**

For this problem, the indices are $\left(S_0 = 4, S_1 = 6, S_{01}^d = -4, D = 8\right)$. For the case without the disposal option, the optimal three-index policy is $(S_0 = 4, S_1 = 6, S_{01} = -6)$. Figure 5.2 displays the costs of different alternatives with and without the disposal option.

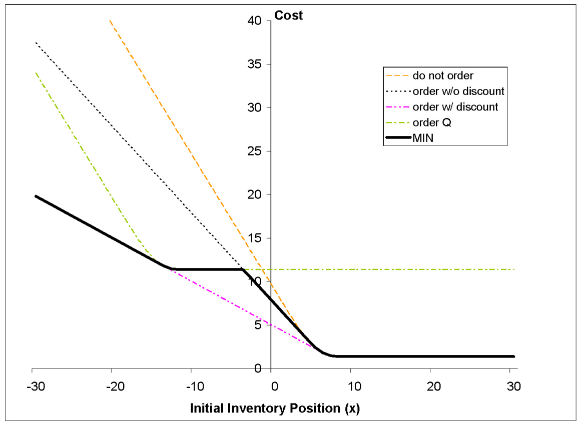
Before we conclude this section, we make two observations: (1) Even for a single period problem we need an additional index to define the optimal policy with disposal. (2) The buyer does not necessarily order and dispose at the beginning of a single period problem.

Infinite Horizon Problem with Inventory Disposal

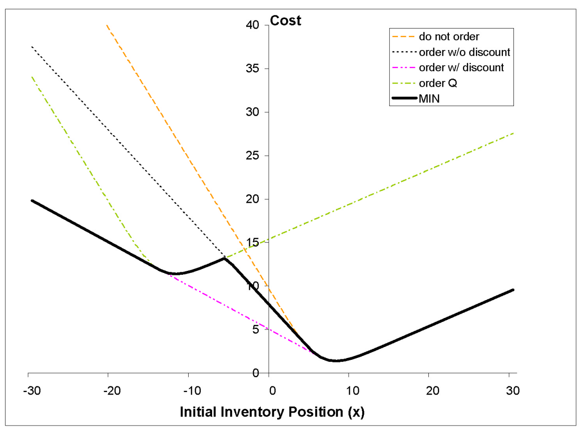
In the infinite horizon, the impact of "irrational characteristics" of quantity discounts diminishes as the buyers may use the item in the future, i.e., the disposal cost of the item increases and the buyer utilizes the disposal option even less. Let us revisit Example 11 and look at the optimal policy for the infinite horizon problem:

Example 12 [Example 1 revisited] *For Poisson*(6)*, with* $s_d = 0$*,* $c_0 = 1.0$*,* $c_1 = 0.5$*,* $p = 0.35$*,* $h = 0.4$*, and* $Q = 20$*, the optimal policy for the infinite horizon problem is as follows:*

[8] We assume that the cost of replenishment of the backordered demand from an alternative source is 1.15 and the penalty cost of customer's waiting for one period is $p = 0.35$. Therefore, $\bar{p} = 1.15 + 0.35 = 1.5$.



(a) With Disposal



(b) Without Disposal

Figure 5.2: The cost functions of different ordering strategies and cost of the optimal policy for Poisson(6), $s_d = 0$, $c_0 = 1.0$, $c_1 = 0.5$, $\bar{p} = 1.5$, $\bar{h} = 0.4$, and $Q = 20$, for various initial inventory positions.

		initial inventory	$\leq$	*-13*	:	*order up to* **7**, *dispose 0*
-12	$\leq$	*initial inventory*	$\leq$	**-7**	:	*order* $Q = 20$, *dispose 0*
-7	$\leq$	*initial inventory*	$\leq$	*14*	:	*order 0, dispose 0*
15	$\leq$	*initial inventory*			:	*order 0, dispose down to* **14**

The buyer disposes inventory only if his inventory position is greater than 15, which is possible only if the buyer starts with that much inventory. His order sizes are not big enough to carry his inventory position to a level greater than 14. Therefore, even though in the single period model the buyer is disposing items, for the same parameter set, in the infinite horizon, disposal is no longer an economic option. Among others, there are two factors that lead to this behavior:

- The buyer does not dispose an item if that item will be used in the next $\frac{c_1}{h}$ periods. Therefore, if the inventory holding cost is low compared to the unit cost, the buyer will prefer to keep an item in the inventory rather than disposing it.
- On average it will take the buyer $\frac{Q}{\mu}$ periods to deplete the inventory that he procures with the discount, where μ is the mean demand. Hence, as long as Q is not too big compared to μ, the buyer will have less incentive to buy with discount and dispose afterwards.

Next, we provide examples where the order quantity is not monotone in initial inventory position even when the disposal option is available. That is, the disposal option does not eliminate *wait-and-see* and *buy-and-hold* strategies of the buyer.

Example 13 *For Poisson*(6)*, with* $s_d = 0$*,* $c_0 = 0.9$*,* $c_1 = 0.7$*,* $p = 0.25$*,* $h = 0.15$*, and* $Q = 30$*, the optimal policy for the infinite horizon problem is as follows:*

					:	
		initial inventory	$\leq$	*-18*	:	*order up to* **12***, dispose 0*
-17	$\leq$	*initial inventory*	$\leq$	**-6**	:	*order* $Q = 30$*, dispose 0*
-5	$\leq$	*initial inventory*	$\leq$	*-1*	:	*order 0, dispose 0*
0	$\leq$	*initial inventory*	$\leq$	*5*	:	*order up to* **5***, dispose 0*
6	$\leq$	*initial inventory*			:	*order 0, dispose down to* **41**

In Example 13, the optimal policy turns out to be a five-index policy: $(S_1^1 = 12, S_{01}^1 = -6, x_l^2 = 0, S_0^2 = 5, D = 41)$. In this example, the buyer follows a *wait-and-see* strategy and there is only one disposal region. The buyer disposes some of his inventory when the inventory position is greater than 41. Note that this is only possible if the buyer has that much inventory to start with; i.e., the buyer does not dispose of inventory in order to be able to order with the discounted price. In this example, the buyer will hold on to an item for at least $\frac{c_1}{h} \approx 4.6$ periods. As demand mean μ is 6 and $Q = 40$, on average it will take him $\frac{Q}{\mu} \approx 6.6$ periods to deplete the inventory that he procures with the discount. Therefore, instead of disposing item, the buyer prefers to backorder some of the demand in order to take advantage of the discounted price.

Example 14 *For Discrete Normal(*$30, \sigma = 4$*) with* $s_d = 0$*,* $c_0 = 1.0$*,* $c_1 = 0.7$*,* $p = 0.45$*,* $h = 0.15$*, and* $Q = 40$*, the optimal policy for the infinite horizon problem is as follows:*

		initial inventory	$\leq$	*-10*	:	*order up to* **30***, dispose 0*
-9	$\leq$	*initial inventory*	$\leq$	**5**	:	*order* $Q = 40$*, dispose 0*
6	$\leq$	*initial inventory*	$\leq$	*17*	:	*order up to* **57***, dispose 0*
18	$\leq$	*initial inventory*	$\leq$	**19**	:	*order* $Q = 40$*, dispose 0*
20	$\leq$	*initial inventory*	$\leq$	*24*	:	*order up to* **24***, dispose 0*
25	$\leq$	*initial inventory*	$\leq$	**169**	:	*order 0, dispose down to* **157**
170	$\leq$	*initial inventory*	$\leq$	*178*	:	*order 0, dispose 0*
179	$\leq$	*initial inventory*			:	*order 0, dispose down to* **179**

In Example 14 the optimal policy turns out to be an eight-index policy: $(S_1^{1:1} = 30, m_u^{1:1} = 5, S_1^{1:2} = 57, S_{01}^{1:2} = 19, S_0^{1:2} = 24, D^1 = 157, d_u^1 = 169, D^2 = 179)$. In this example, the buyer follows a *buy-and-hold* strategy and there are *two* disposal regions. Again disposal is an option only if the buyer starts with a high initial inventory.

Examples 13 and 14 show that the disposal option does not eliminate the regions that cause order quantities to lose monotonicity in the original problem, and we continue to observe the same effects in the revised model as we do in the original model. In these examples, the disposal option complicates the policy definition by increasing the number of indices. However, the buyer uses disposal only if he starts with a high inventory position. Therefore, even if he disposes items, he writes-off inventory only once at the beginning of the planning horizon. The inventory never falls into the disposal region again after the first period since we assume stationary demand. Disposal is a more practical option when demand is nonstationary due to some seasonality. In this case, the initial inventory position may fall into the disposal region when mean demand shifts down.

Next, we provide an example where the buyer disposes inventory many times. However, disposal option still does not provide monotonicity of the order quantities and even complicates the policy further by adding an additional index. In this example, $\frac{Q}{\mu}$ is greater than 7 which is not very practical and makes holding inventory very costly after placing an order of Q. We can create the same effect by providing a very high h.

Example 15 *For Poisson(6) with* $s_d = 0$, $c_0 = 1.0$, $c_1 = 0.5$, $p = 0.25$, $h = 0.4$, *and* $Q = 45$, *the optimal policy for the infinite horizon problem is as follows:*

		initial inventory	$\leq$	*-34*	:	*order up to* **11**, *dispose 0*
-33	$\leq$	*initial inventory*	$\leq$	*-26*	:	*order* $Q = 45$, *dispose 0*
-25	$\leq$	*initial inventory*	$\leq$	**-20**	:	*order* $Q = 45$, *dispose down to* **19**
-19	$\leq$	*initial inventory*	$\leq$	*-3*	:	*order 0, dispose 0*
-2	$\leq$	*initial inventory*	$\leq$	*5*	:	*order up to* **5**, *dispose 0*
6	$\leq$	*initial inventory*	$\leq$	*19*	:	*order 0, dispose 0*
20	$\leq$	*initial inventory*			:	*order 0, dispose down to* **19**

The optimal policy turns out to be a five-index policy: $(S_1^1 = 11, S_{01}^1 = -20, x_l^2 = -2, S_0^2 = 5, D = 19)$. In this example, there are two ordering regions and two disposal regions. In the first disposal region, the buyer orders Q and dispose some of the inventory. In the second disposal region, the buyer disposes only if he starts the planning horizon with a high initial inventory. Notice that, the buyer continues to use the wait-and-see strategy even when disposal option is available.

Note that for the parameter sets where the buyer follows a buy-and-hold strategy, it is quite difficult to generate an example where the buyer procures with the discount and then disposes items. The inventory holding cost should be quite low in order to tempt the buyer to engage in buy-and-hold strategy in the first place. However, in that situation, the buyer does not have an incentive to dispose. Therefore, the buy-and-hold strategy is not affected much by the disposal option (Example 14). We continue to observe parameter sets where the buyer prefers to hold inventory in order to take advantage of the discounted price even when disposal is an option.

5.2 Appendix for Chapter 4

5.2.1 The conditional distribution of X_1

Proposition 4 *Let X be distributed as $N(\mu, \Sigma)$ and*

$$X = \begin{vmatrix} X_1 \\ X_2 \end{vmatrix} \quad \Sigma = \begin{vmatrix} \Sigma_{11} & \Sigma_{12} \\ \Sigma_{21} & \Sigma_{22} \end{vmatrix} \quad \mu = \begin{vmatrix} \mu_1 \\ \mu_2 \end{vmatrix}$$

Then the conditional distribution of X_1, given X_2 is $N(\mu_1 + \Sigma_{12}\Sigma_{22}^{-1}(X_2 - \mu_2), \Sigma_{11} - \Sigma_{12}\Sigma_{22}^{-1}\Sigma_{21})$.

Proof: Available in Johnson and Wichern (1992). □

5.2.2 Matrix Form of the problem

Before we start looking at the supplier's problem, we introduce some concepts about multivariate normal distributions which helps to derive some useful results. We have X and $\hat{X}$ as follows

$$X = \begin{vmatrix} X_1 \\ X_2 \\ \vdots \\ X_{T-1} \\ X_T \end{vmatrix} \quad \hat{X} = \begin{vmatrix} \hat{X}_1 \\ \hat{X}_2 \\ \vdots \\ \hat{X}_{T-1} \\ \hat{X}_T \end{vmatrix} = \begin{vmatrix} \hat{X}_1(1) & \hat{X}_1(2) & \cdots & \hat{X}_1(K-1) & \hat{X}_1(K) \\ \hat{X}_2(1) & \hat{X}_2(2) & \cdots & \hat{X}_2(K-1) & \hat{X}_2(K) \\ \vdots & \vdots & \ddots & \vdots & \vdots \\ \hat{X}_{T-1}(1) & \hat{X}_{T-1}(2) & \cdots & \hat{X}_{T-1}(K-1) & \hat{X}_{T-1}(K) \\ \hat{X}_T(1) & \hat{X}_T(2) & \cdots & \hat{X}_T(K-1) & \hat{X}_T(K) \end{vmatrix}$$

If we put the problem in matrix notation, we have the following model

$$\begin{aligned} (1-\phi B)X &= (1-\theta B)\epsilon + A_1 \\ X &= (1-\phi B)^{-1}(1-\theta B)\epsilon + (1-\phi B)^{-1}A_1 \end{aligned}$$

$$(1-\phi B) = \begin{vmatrix} 1 & 0 & 0 & \cdots & 0 & 0 & 0 \\ -\phi & 1 & 0 & \cdots & 0 & 0 & 0 \\ 0 & -\phi & 1 & \cdots & 0 & 0 & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots & \vdots & \vdots \\ 0 & 0 & 0 & \cdots & 1 & 0 & 0 \\ 0 & 0 & 0 & \cdots & -\phi & 1 & 0 \\ 0 & 0 & 0 & \cdots & 0 & -\phi & 1 \end{vmatrix} \quad (1-\theta B) = \begin{vmatrix} 1 & 0 & 0 & \cdots & 0 & 0 & 0 \\ -\theta & 1 & 0 & \cdots & 0 & 0 & 0 \\ 0 & -\theta & 1 & \cdots & 0 & 0 & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots & \vdots & \vdots \\ 0 & 0 & 0 & \cdots & 1 & 0 & 0 \\ 0 & 0 & 0 & \cdots & -\theta & 1 & 0 \\ 0 & 0 & 0 & \cdots & 0 & -\theta & 1 \end{vmatrix}$$

$$A_1 = \begin{vmatrix} (1-\phi)\mu + \phi X_0 - \theta\epsilon_0 \\ (1-\phi)\mu \\ \vdots \\ (1-\phi)\mu \end{vmatrix} \quad (1-\phi B)^{-1} = \begin{vmatrix} 1 & 0 & 0 & \cdots & 0 & 0 & 0 \\ \phi & 1 & 0 & \cdots & 0 & 0 & 0 \\ \phi^2 & \phi & 1 & \cdots & 0 & 0 & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots & \vdots & \vdots \\ \phi^{T-3} & \phi^{T-4} & \phi^{T-5} & \cdots & 1 & 0 & 0 \\ \phi^{T-2} & \phi^{T-3} & \phi^{T-4} & \cdots & \phi & 1 & 0 \\ \phi^{T-1} & \phi^{T-2} & \phi^{T-3} & \cdots & \phi^2 & \phi & 1 \end{vmatrix}$$

Bibliography

[1] R. Anupindi and R. Akella. An inventory model with commitments. *Northwestern University, Working Paper*, 1993.

[2] R. Anupindi and Y. Bassok. Supply contracts with quantity commitments and stochastic demand. In S. Tayur, R. Ganeshan, and M. Magazine, editors, *Quantitative Models for Supply Chain Management*, pages 197–232. Kluwer Academic Publishers, Boston, MA, 1998.

[3] J. S. Armstrong. Extrapolation for time-series and cross-sectional data. In J. S. Armstrong, editor, *Principles of Forecasting:A Handbook for Researchers and Practitioners*. Kluwer Academic Publishers, 2001.

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

[5] Y. Aviv. Gaining benefits from joint forecasting and replenishment processes: The case of auto-correlated demand. *Manufacturing and Service Operations Management*, 4:55–74, 2002.

[6] Y. Aviv. A time-series framework for supply-chain inventory management. *Operations Research*, 51(2):210–227, 2003.

[7] K. S. Azoury. Bayes solution to dynamic inventory models under unknown demand distribution. *Management Science*, 31:1150–1160, 1985.

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

[9] Roy Batchelor and Pami Dua. Forecaster diversity and the benefits of combining forecasts. *Management Science*, 41(1):68–75, 1995.

[10] W.C. Benton and S. Park. A classification of literature on determining the lot size under quantity discounts. *European Journal of Operations Research*, 92:219–238, 1996.

[11] D.P. Bertsekas. *Dynamic Programming and Optimal Control*. Athena Scientific, Belmont, Massachusetts, 1995.

[12] G. E. P. Box and G. M. Jenkins. *Time Series Analysis: Forecasting and Control*. San Francisco, Holden-Day, 1970.

[13] J.M. Buchanan. The theory of monopolistic quantity discounts. *Review of Economic Studies*, 20(3):199–208, 1953.

[14] G. Cachon. Managing supply chain demand variability with scheduled ordering policies. *Management Science*, 45(6):843–856, 1999.

[15] G. Cachon and S. Netessine. Game theory in supply chain analysis. In S. D. Wu D. Simchi-Levi and Z. Shen, editors, *Handbook of Quantitative Supply Chain Analysis: Modeling in the eBusiness Era*. Kluwer, 2004.

[16] G. P. Cachon and M. Lariviere. Capacity choice and allocation: Strategic behavior and supply chain performance. *Management Science*, 45:1091–1108, 1999a.

[17] G. P. Cachon and M. Lariviere. An equilibrium analysis of linear, proportional and uniform allocation of scarce capacity. *IIE Transactions*, 45:835–849, 1999b.

[18] F. Chen, A. Federgruen, and Y. Zheng. Coordination mechanism for a distribution system with one supplier and multiple retailers. *Management Science*, 47(5):693–708, 2001.

[19] F. Chen and R. Samroengraja. Order volatility and supply chain costs. *Columbia University, Working Paper*, 2003.

[20] K.L. Cheung and H.L. Lee. The inventory benefits of shipment coordination and stock rebalancing in a supply chain. *Management Science*, 48(2):300–306, 2002.

[21] C.J. Corbett and X. de Groote. A supplier's optimal quantity discount policy under asymmetric information. *Management Science*, 46(3):444–500, 2000.

[22] R.J. Dolan. Quantity discounts: Managerial issues and research opportunities. *Marketing Science*, 6(1):1–23, 1987.

[23] F. Erhun and S. Tayur. Enterprise-wide optimization of total landed cost at a grocery retailer. *Operations Research*, 51(3):343–353, 2003.

[24] A. Garbor. A note on block tariffs. *Review of Economic Studies*, 23:32–41, 1955.

[25] J. Geweke and C. Whiteman. Bayesian forecasting. In C.W.J. Granger G. Elliott and A. Timmermann, editors, *The Handbook of Economic Forecasting*. Elsevier, 2006.

[26] S. C. Graves. A tactical planning model for a job shop. *Operations Research*, 34:522–533, 1986a.

[27] S. C. Graves, D. B. Kletter, and W. B. Hetzel. A dynamic model for requirements planning with application to supply chain optimization. *Operations Research*, 46:S35–S49, 1998.

[28] S. C. Graves, D. B. Kletter, and W. B. Hetzel. A single-item inventory model for a non-stationary demand process. *Manufacturing and Service Operations Management*, 1(1):50–61, 1999.

[29] S. C. Graves, H.C. Meal, S. Dasu, and Y. Qiu. Two stage production planning in a dynamic environment. In S. Axsater, C. Schneeweiss, and E.Silver, editors, *Multi-Stage Production Planning and Control: Lecture Notes in Economics and Mathematical Systems*, pages 9–43. Springer-Verlag, Berlin, 1986b.

[30] Banta Publications Group. Transportation costs continue to rise. *Banta Mailbox*, December 2004.

[31] R. Gullu. On the value of information in dynamic production/inventory problems under forecast evolution. *Naval Research Logistics*, 43:289–303, 1986.

[32] W. H. Hausman. Sequential decision problems: A model to exploit existing forecasters. *Management Science*, 16:B93–B111, 1969.

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

[34] Industry Directions Inc. Manufacturers leverage supply chain strategy to offset rising energy costs. *Executive Brief, www.industrydirections.com*, October 2005.

[35] A.P. Jeuland and S.M. Shugan. Managing channel profits. *Marketing Science*, 2(3):239–272, 1983.

[36] R. A. Johnson and D. W. Wichern. *Applied Multivariate Statistics*. Prentice Hall, 1992.

[37] J.V. Jucker and M.J. Rosenblatt. Single-period inventory models with demand uncertainity and quantity discounts: Behaviorial implications and a new solution procedure. *Naval Research Logistics*, 32:537–550, 1985.

[38] T. Kohonen. *Self Organizing Maps*. Springer, 2001.

[39] R. Lal and R. Staelin. An approach for developing an optimal discount policy. *Management Science*, 30(12):1524–1539, 1984.

[40] M. A. Lariviere and E. L. Porteus. Stalking information: Bayesian inventory management with unobserved lost sales. *Management Science*, 45:346–363, 1999.

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

[42] H.L. Lee and M.J. Rosenblatt. A generalized quantity discount pricing model to increase supplier's profits. *Management Science*, 32(9):1177–1185, 1986.

[43] K. Moinzadeh and S. Nahmias. Adjustment strategies for a fixed delivery contract. *Operations Research*, 48(3):408–423, 2000.

[44] J.P. Monahan. A quantity discount pricing model to increase vendor profits. *Management Science*, 30(6):720–726, 1984.

[45] C.L. Munson and M.J. Rosenblatt. Theories and realities of quantity discounts: An exploratory study. *Production and Operations Management*, 7(4):352–369, 1998.

[46] R. M. Neal. Slice sampling. *The Annals of Statistics*, 31(3):705–767, 2003.

[47] E.L. Porteus. On the optimality of generalized (s, S) policies. *Management Science*, 17:411–426, 1971.

[48] J. W. Sammon. A nonlinear mapping for data structure analysis. *IEEE Transactions on Computers*, C-18:401–409, 1969.

[49] H. Scarf. Bayes solutions of the statistical inventory problem. *Annals of Mathematical Statistics*, 30:490–508, 1959.

[50] P.S. Sethi. A quantity discount lot size model with disposals. *International Journal of Production Research*, 22:31–39, 1984.

[51] Adesso Solutions. A double-edged sword. *Business Credit*, 107(9):56–59, 2005.

[52] C. Terwiesch, J. Z. Ren, T. H. Ho, and M. Cohen. An empirical analysis of forecast sharing in the semiconductor equipment industry. Working paper, Wharton Business School, University of Pennsylvania, Philadelphia, PA, 2003.

[53] B. Toktay and L. M. Wein. Analysis of a forecasting-production-inventory system with stationary demand. *Management Science*, 47:1268–1281, 2001.

[54] A. Tsay, S. Nahmias, and N. Agrawal. Modeling supply chain contracts: A review. In S. Tayur, R. Ganeshan, and M. Magazine, editors, *Quantitative Models for Supply Chain Management*, pages 301–336. Kluwer Academic Publishers, Boston, MA, 1998.

[55] Z.K. Weng. Channel coordination and quantity discounts. *Management Science*, 41:1509–1522, 1995.