PLANNING MODELS FOR PRICE PROMOTIONS IN MULTI-LEVEL SUPPLY CHAINS

By

YIQIANG SU

A DISSERTATION PRESENTED TO THE GRADUATE SCHOOL
OF THE UNIVERSITY OF FLORIDA IN PARTIAL FULFILLMENT
OF THE REQUIREMENTS FOR THE DEGREE OF
DOCTOR OF PHILOSOPHY

UNIVERSITY OF FLORIDA

2013

I dedicate this to my girlfriend, Zhuofei for being there for me throughout the entire doctorate program.

ACKNOWLEDGMENTS

First and foremost, to my parents, your love, support and understanding have been invaluable. Then I would like to express the deepest appreciation to my committee chair, Professor Joseph Geunes, who has great impact on my academic life. I thank you for your enlightening ideas, extensive knowledge about supply chain management and remarkable patience. Without your guidance and help this dissertation would not have been possible. I also wish to thank my committee members, Professor Janice Carrillo, Professor Cole Smith, and Professor Yongpei Guan who were more than generous with their expertise and precious time. I thank you all for suggestions on my dissertation. Finally, I take this opportunity to record my sincere thanks to all the faculty members of the Industrial and Systems Engineering program at the University of Florida, and my graduate student colleagues. Special thanks to Ruiwei, Qianfan, Siqian, Zhili, Chin Hon, Kelly, Hongsheng, Andrew, Soheil, Cinthia, Prince, and Clay, your friendship made my doctorate study a memorable experience of my life.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

Abstract of Dissertation Presented to the Graduate School
of the University of Florida in Partial Fulfillment of the
Requirements for the Degree of Doctor of Philosophy

PLANNING MODELS FOR PRICE PROMOTIONS IN MULTI-LEVEL SUPPLY CHAINS

By

Yiqiang Su

December 2013

Chair: Joseph Geunes
Major: Industrial and Systems Engineering

Price promotions are playing a significant role in the retail industry; however, inefficiency of price promotions has been demonstrated both in theoretical works and industrial practice. As a result, other price strategies, like every day low price (EDLP) strategies, are recommended to replace price promotions. Since there is plenty of evidence that price promotions are not replaceable from the consumer's perspective, price promotions are certain to continue existing in the near future. This dissertation is going to demonstrate a number of profitable price promotion scenarios, and develop tools to help supply chain participants to increase price promotion performance and resolve potential conflicts that may occur in price promotion activities.

We first consider a deterministic, two-stage supply chain model composed of a single supplier and a single retailer. In this model, the supplier periodically offers the retailer a price discount at a single time point within each "planning cycle." The retailer needs to determine its order and pricing policy in order to maximize its profit. In response to the retailer's policy, the supplier must determine its transportation and inventory policy in order to minimize its total operations cost. We demonstrate that even though the use of trade promotions can indeed increase a retailer's and supplier's operations costs, these costs may be more than offset by increased revenues, even in the absence of explicit coordination. We provide a broad set of computational results that validate this conclusion and discuss the resulting managerial insights.

Then we consider a stochastic, two-stage supply chain model in which a supplier serves a set of stores in a retail chain. In this model, the supplier offers periodic off-invoice price discounts to the retail chain over a finite planning horizon under uncertainty in demand. Based on the price discounts offered by the supplier, and after store demand uncertainty is resolved, the retail chain determines individual store order quantities in each period. Because the supplier offers store-specific prices, the retailer may ship inventory between stores, a practice known as diverting. We demonstrate that, despite the resulting bullwhip effect and associated costs, a carefully designed price promotion scheme can improve the supplier's profit when compared to the case of everyday low pricing (EDLP). We model this problem as a stochastic bilevel optimization problem with a bilinear objective at each level and with linear constraints. We provide an exact solution method based on a Reformulation-Linearization Technique (RLT). In addition, we compare our solution approach with a widely used heuristic and another exact solution method developed by Al-Khayyal [7] in order to benchmark its quality.

Finally, we consider an optimization problem that helps a clicks-and-mortar retailer coordinate product assortments, retail prices and inventories between a physical channel and an online channel. In this problem, we consider a clicks-and-mortar retailer who sells products to end-consumers through both physical and virtual channels. We use a multinomial logit (MNL) model to describe the consumer purchase process and a chance constraint to limit the probability of shortage. We model this problem as a chance-constrained and two-stage stochastic problem. We adopt an effective combined sample average approximation algorithm developed by Wang and Guan [97] to solve this problem. In addition, we also propose a greedy and local search heuristic as an alternative solution approach for large size problems when computing power is limited or solution time is critical.

CHAPTER 1
INTRODUCTION

Nowadays, in an increasingly competitive retailing environment, price promotions are becoming important marketing activities which affect every stage in a supply chain. An A.C. Nielson survey shows that about 41% of consumers actively look deals at retail stores [72]. Another A.C. Nielson survey illustrates that a large percentage (12% - 25%) of European retailer sales are made via sales promotions [51]. Moreover, the use of trade promotions by suppliers of consumer packaged goods to distributors rose from $8 billion in 1990 to approximately $80 billion in 2004 [52].

In this dissertation, we will to consider two distinct types of price promotions in supply chains: retail sales promotions and trade promotions. Retail sales promotions target consumers, while trade promotions are offered by an upstream supply chain player to a downstream player. We stress the difference between these two kinds of sales promotions because they affect the supply chain players and consumers in different ways.

Despite being heavily used in the retail industry, price promotions are not always effective for either suppliers or retailers. With the lion's share (52%) [6] of the money spent on advertising and promotions going to trade rather than sales promotions, 85% of suppliers believe that their trade promotion dollars are not being spent effectively [1], and only 19% think they get a good value for their money [19]. One of the major reasons for the ineffectiveness of the trade promotion is the adoption of "off-invoice" deals. In an "off-invoice" trade promotion, a supplier simply deducts some percentage from the invoice price for purchases over a set of periods. Thus, suppliers have no control over whether retailers will actually pass their promotional prices on to end consumers. In fact, a survey shows that, on average, retailers only pass 60% of trade funds to consumers [23]. Besides this low pass-through rate, there are two retailer strategies that reduce the effectiveness of trade promotions. First, retailers may respond to

an off-invoice trade promotion by engaging in *forward-buying*, i.e., the retailer takes advantage of the supplier's temporary discount by purchasing a quantity that far exceeds its current needs, and subsequently only applies a discount to part of this order when selling to consumers. When the promotion period expires, the retailer sells the remaining inventory that it purchased at a discount to its consumers at the regular price. Second, retailers may take advantage of off-invoice trade promotions by engaging in *diverting*, i.e., ordering from one store and redirecting the shipment to another store. Bronnenberg, Dhar, and Dubé [29] show that geographic variation in brand shares for leading consumer brands is both substantial and persistent. Consequently, the marginal effectiveness of promotions and prices varies considerably among retail stores from around the country [28]. To accommodate geographic differences in sales of similar products, suppliers may offer certain off-invoice trade promotions only in certain regions of the country. Large national retail chains may circumvent these restrictions by purchasing a product at a lower wholesale price where the deal is offered, and subsequently shipping the item to stores in other regions (where the deal is not offered) for sale at the regular price. Conventional wisdom holds that 5% to 10% of grocery products on promotion are diverted [22].

These forward-buying and diverting strategies can both result in the bullwhip effect, which is viewed in extremely negative terms because of its negative impacts on supply chain operations costs. Lee et al. [64, 65] also characterized price variations as one of the major causes of the bullwhip effect.

For a retailer, although sales promotions can increase sales volume significantly in the short term, some researchers showed that a large portion of this immediate increase in sales comes from brand switching [26]. It is possible that when a retailer offers a discount for a particular product, some consumers choose to buy the discounted product instead of their first-choice product with a higher price. This so-called *cannibalization* effect can potentially make the sales promotions unprofitable even in the short term.

Moreover, some researchers [26] argue that sales promotions have either no long term effect or negative long term effect. This occurs because frequent sales promotions increase price sensitivity and undermine brand loyalty [34].

Because price promotions are not extremely efficient, Lee et al. [64, 65] suggested that suppliers and distributors adopt corresponding management practices (an EDLP strategy, for example) to stabilize prices. Meanwhile, some suppliers, like Zara® and P&G®, have already begun adopting EDLP strategies and achieved business successes. However, we believe in the near future price promotions will continue playing an important role in the retail industry. First, trade promotions allow supplier to pass temporary reduced raw material costs to retailers and final consumers, which may increase a supplier's sales and enhance a supplier's capacity utilization. Second, for some consumers, price promotions cannot be replaced by EDLP strategies. EDLP strategies can minimize both purchasing cost for the consumers and operations costs for the retailers and suppliers; however, sales promotions can provide consumers with more benefits beyond merely saving money [34]. For example, sales promotions can enhance a consumer's self-perception as a savvy shopper. As long as consumers continue responding to price promotions, retailers and suppliers should keep spending on price promotions. A 2012 AMG Strategic Advisors study backs up this idea, and it shows that out of more than 768 suppliers, 56% of them place high or higher priority on increasing spending on trade promotion funds in the next three years. A 2012 IEG report also shows that in the last three years sales promotion spending grew 2% per year on average, and it predicts that this spending will increase by 3% in 2013. Since price promotions are currently not replaceable, models that improve the effectiveness and efficiency of price promotions should be developed.

CHAPTER 2
PRICE PROMOTIONS, OPERATIONS COST, AND PROFIT IN A TWO-STAGE SUPPLY CHAIN

2.1 Problem Description and Literature Survey

The phenomenon in which demand variability increases as one moves upstream in the supply chain has been often observed in practice. This so called "bullwhip effect" has drawn a surge of research interest for half a century. Forrester [47] first identified the phenomenon of increasing demand variability as one moves up a supply chain via a series of case studies. He suggested that this effect was caused by time varying behaviors within industrial organizations. Sterman [90] used the well-known "Beer Distribution Game" to show that the amplitude and variance of orders increase steadily from consumer to retailer to factory, attributing this phenomenon to players' "misperceptions of feedback."

Previous research has mainly focused on identifying causes of the bullwhip effect, quantifying the cost impacts of the bullwhip effect, and developing strategies to reduce the bullwhip effect. Arguably the most influential work in recent years was provided by Lee, Padmanabhan, and Whang [64, 65]. They showed that even under the assumption that the supply chain members are rational, optimizing agents, the so-called "bullwhip effect" may still exist, and they identify four main causes of this effect: the use of certain demand forecasting methods, supply shortage games, order batching, and price fluctuations. In addition, they also proposed mechanisms to counter the negative impacts of the bullwhip effect on operations costs, one of which was the use of an EDLP strategy to counter the component of the bullwhip effect that results from price fluctuations. Chen, Drezner and Simchi-Levi [35, 36] quantified the impact of demand forecasting on the bullwhip effect for a two-echelon supply chain consisting of a single supplier and a single retailer. In their study, the demand distribution parameters are not known with certainty, and the retailer therefore uses moving average or exponential smoothing forecasts to estimate the demand mean and variance. These estimates are

then used to implement an order-up-to inventory replenishment policy. In measuring the bullwhip effect, Fransoo and Wouters [48] discussed potential measurement problems as a result of data aggregation, incompleteness of data, and the isolation of demand data. They used the ratio of the coefficient of variation of orders placed to the coefficient of variation of orders received as a measure of the bullwhip effect. Cachon [31] also used this approach to measure the bullwhip effect. He investigated a supply chain with one supplier and N retailers facing stochastic demand, who implement scheduled ordering policies, and found that the supplier's demand variability declines as the retailer's order intervals increase or the batch size increases. Potter and Disney [81] focused on the impact of batch size on the bullwhip effect, and they used the ratio of variance of orders placed to the variance of orders received to measure the bullwhip effect.

However, little model-based research has addressed the extent to which the bullwhip effect might be a necessary evil in a profit-maximizing supply chain. That is, if an activity simultaneously increases revenue and cost (as a result of the bullwhip effect), how does a decision maker choose the level of this activity that optimizes system performance? In our study, this activity corresponds to a price promotion. As we mentioned in Chapter 1, price promotions still prevail in industrial practices nowadays, and we believe there are many reasons that price promotions will not be completely replaced by EDLP strategies in the near future. Research in marketing science also provides evidence that well-designed price variations can benefit the entire supply chain system.

Blattberg, Briesch, and Fox [24] summarized this view by contending that: 1) temporary retailer price reductions substantially increase sales; 2) the frequency of deals changes a consumer's reference price; and 3) the greater the frequency of deals, the lower the height of the deal spike. Blattberg, Eppen, and Lieberman [25] showed that food retailers predominantly prefer offering temporary price discounts to adopting

an EDLP strategy, attributing this to a desire to increase market share, and to attract newcomers to their products by offering a lower risk level. Hoch, Dreze, and Purk [57] thoroughly investigated the use of an EDLP strategy, and found that this strategy tends to improve a supplier's profit very little, while potentially leading to big losses for a retailer. Ailawadi, Farris and Shames [6] used a numerical example to demonstrate that a well-designed trade promotion can increase the total supply chain system profits and that the upstream player gains a larger share of the total profit than the case in which the supplier fixes a single price and does not use promotions. In their single-supplier, single-retailer study, the supplier's unit costs are fixed, while the retailer's unit cost equals a fixed value plus the supplier's wholesale price. Moreover, their model does not account for fixed order costs or inventory holding costs, and therefore the bullwhip effect is not a factor. Chandon, Wansink, and Laurent [34] built a framework addressing the multiple consumer benefits of sales promotions, and they classified these benefits into six categories: monetary savings, quality increases, convenience, value expression, exploration, and entertainment. They also recommended against a retailer's blind use of EDLP, arguing that consumers respond to sales promotions for reasons that extend beyond monetary benefits.

The presence of apparent inconsistencies between the marketing and operations views of promotions leads us to wonder how price fluctuations affect total system profit in a supply chain. Since the conclusions from the marketing science literature typically do not explicitly consider supply chain operations costs, a mathematical model that accounts for these operations costs in addition to the demand-side effects of price promotions may serve to partially reconcile these conflicting views.

Some past research has partially addressed this question, taking a retailer's view. One group of studies is under the assumption of deterministic demand. Ardalan [14, 15] developed a model that determines both the retailer's optimal price and ordering policies in response to a one-time only price discount using a general price-demand relationship.

Arcelus and Srinivasan [12] generalized this work by accounting for potential forward buying at the retailer. For the more general case in which price discounts can be offered on more than one occasion, Sogomonian and Tang [88] developed a model that determines optimal promotion and production decisions for a single profit-maximizing firm using a mixed-integer program. Based on the assumptions that promotion levels belong to a finite set and the consumer response to promotions depends on the time elapsed since the last promotion, they proposed an approach to solve the problem globally by reformulating their problem as a "longest path" problem on a network.

Another group of studies related to sales promotions is under the assumption of stochastic demand. Petruzzi and Dada [80] considered a price-setting firm that stocks a single product subject to random, price-dependent demand, with the objective of determining stock levels over multiple periods. They developed a condition under which the solution to this problem is stationary and myopic. One limitation of their research is the imposition of a constant price assumption. For a joint pricing and inventory management problem with stochastic demand (in which prices may change dynamically over time), Thowsen [92] showed that the optimal pricing/inventory policy is a base-stock list-price policy under the assumptions that: i) the expected demand curve and stockout costs are linear functions; ii) holding costs are convex; and iii) the density function of the random component of consumer demand is PF_2 (a Pólya frequency function of order 2). Cheng and Sethi [38] used a Markov decision process (MDP) model for a joint inventory and promotion decision problem in which the promotion has only two states (on and off). Under certain conditions, they showed that the optimal promotion and ordering policy is an (S^0, S^1, P) policy. That is, if the initial inventory level is at least P, the product is promoted; if the initial inventory is less than S^1 and the product is promoted, an order is placed to increase the inventory position to S^1; if the initial inventory is less than S^0 and if the product is not promoted, an order is placed to increase the inventory position to S^0.

Beyond the retailer's inventory and pricing policies, we are also interested in how the systemwide (supplier and retailer) policies interact to drive performance. To this end, Goyal [53] proposed an integrated inventory model for a two-echelon system; his work compared independent and joint approaches for a system inventory control problem. Ernst and Pyke [46] examined inventory policies at both a warehouse and the retailer it serves, and they also consider the warehouse's optimal shipping policy. They characterized the impact of consumer demand variability on optimal warehouse trucking capacity. In each of these settings, price was treated as a fixed (and stationary) parameter over the planning horizon.

In this chapter, we provide a mathematical model of a two-echelon supply chain in which the supplier's and retailer's prices may be periodically reduced in conjunction with a promotion. Since we wish to focus on pricing effects, we assume that consumer demand at the retail echelon is deterministic and price-dependent. This assumption permits us to gain insight into such problems by isolating the way in which pricing policies influence operations costs in the absence of random demand. Our goal is to capture the way in which both a retailer and its consumers react to promotions, as well as how this, in turn, translates to the bullwhip effect and impacts operations costs and profit levels.

Gavirneni [50] considered a model that is similar in spirit to ours, although it uses a different set of assumptions and factors that drive performance. In particular, he studied a capacitated two-echelon supply chain with stochastic demand. He showed that if information sharing occurs in the supply chain, price fluctuations can improve total system performance. His study considered a trade promotion only and did not allow for the retailer to pass any of the discount to consumers. In contrast, in our model, the impact of retailer price discounts on demand serves as the main driver of the increased system profit resulting from promotions.

The remainder of this chapter is organized as follows. In Section 2.2 we describe our supply chain model. Subsection 2.2.1 states our assumptions about consumer demand, followed by Subsections 2.2.2 and 2.2.3, which derive mathematical models to optimize retailer and supplier profits, respectively. In Subsection 2.2.4, we elaborate on how we compute the variance of retailer orders and the variance of consumer demand, which together permit characterizing the bullwhip effect. Section 2.3 illustrates how promotions affect profits in our model via a set of numerical tests; we compare the retailer's profit, the supplier's profit, and the total system profit in both the promotion scenario and the constant-price scenario.

2.2 Problem Statement and Formulation

We begin by considering a simple distribution system structure, namely, a single supplier who sells a product to a retailer, who in turn sells it to consumers. We assume that consumer demand occurs at a constant, deterministic rate that is known by the retailer and is also price-dependent.

There is one major reason why supplier price fluctuations can lead to the bullwhip effect: the retailer may respond to a temporary price discount by engaging in forward buying, i.e., the retailer takes advantage of the supplier's temporary discount by purchasing a quantity that far exceeds its current needs, and subsequently only applies a discount to part of this order when selling to consumers. When the promotion period expires, the retailer sells the remaining inventory that it purchased at a discount from the supplier to its consumers at the regular price. This makes the retailer's ordering pattern differ from the consumer demand pattern, and consequently, the variation in a retailer's ordering quantity may be much larger than the variation in end-consumer demand.

Clearly, if the supplier and the retailer work cooperatively, this can enable the supplier to access the retailer's point-of-sale data, which makes it possible for the supplier to offer discounts only on the retailer's "sell-through" units (i.e., units sold by the retailer at a reduced price to consumers). As a result, the retailer can no longer adopt

a forward buying policy, and the driver disappears. However, countering the bullwhip effect is not the emphasis of our work, as we consider systems in which such explicit coordination is absent, and investigate the relationship between price discounts and the bullwhip effect in such cases. That is, in this chapter we seek to develop and explore a model that validates the coexistence of the bullwhip effect and increased profits when compared to an EDLP strategy.

We consider a situation in which the supplier periodically offers the retailer a price discount at a single time point within each "planning cycle" (where the planning cycle duration is defined by the time between promotions or price discounts). The retailer is not required to pass this price reduction on to its consumers, since the revenue from the extra demand generated by the lower price may not offset the corresponding cost. The retailer thus needs to consider three factors: (i) the quantity to be ordered at a discount from the supplier; (ii) the discounted price to be offered to its consumers; and (iii) the forward buy quantity, i.e., the quantity it will purchase at a discount in excess of its normal order quantity that it will not "sell-through" at a discount. After the retailer sets its ordering policy, the retailer's order and pricing information is passed to the supplier without delay. In response to the retailer's ordering and pricing policy, the supplier must determine its transportation and inventory policy in order to minimize its total operations cost.

We compare two scenarios, denoted by S_0 and S_1. The first scenario, S_0, is the normal scenario, under which the supplier's price is constant, and the retailer sells the product at a price P_0 per unit, with a corresponding consumer demand rate of D_0. Under this scenario, the retailer repeatedly orders in lots of size q_0 when its inventory is depleted, as in the normal economic order quantity (EOQ) model. This so-called normal scenario will correspond to the EDLP case. The second scenario, S_1, represents the situation in which the supplier offers a discount to the retailer at the beginning of each planning cycle. Based on previous research [76, 91], if the retailer is a rational and

optimizing agent, the retailer will accept this offer and place a special order of size Q_1. The retailer will then offer a (possibly) discounted price of $P_1 \leq P_0$ to its consumers. It is a common practice in retailing industry to "highlight" the promotional offers in certain ways to draw consumer attention, and, in this chapter, we also assume whenever the retailer offers a price discount, the retailer displays this promotional offer effectively, so the consumer can distinguish the difference between promotional price and regular price without difficulty. The cost associated with the "highlight" or "display" activity is usually negligible compared to other costs, for example, a common practice to feature a price discount is to change the color of price tag. A retail price discount with in-store retail display induces a consumer demand rate of $D_1 \geq D_0$. We assume the retailer subsequently returns to its normal price P_0 and order quantity q_0 after depleting all of the special order of size Q_1. In both scenarios, we assume that no shortages are permitted at the retailer.

We are interested in capturing the two primary metrics of total system profit and the bullwhip effect. The total system profit is equal to the sum of the retailer's and the supplier's profit, where the retailer's profit is defined as the difference between its revenue and all costs, including variable purchasing cost, inventory holding cost, and fixed order cost; similarly, the supplier's profit equals its revenue less its production, inventory holding, setup, and transportation costs. We discuss our computation of the bullwhip effect later in Section 2.2.4.

2.2.1 Impact of Promotions on Demand

The notation we use is summarized below:

- d : supplier's unit discount (in dollars) to the retailer;
- δ : retailer's unit discount (in dollars) to its own consumers;
- i : retailer's inventory holding cost rate per unit per year;
- c : retailer's non-discounted purchase price;

- A : retailer's fixed order cost;
- S_0 : scenario in which $d = 0$;
- S_1 : scenario in which $d > 0$;
- P_s : retailer's selling price for alternative $s = 0, 1$;
- D_s : retailer's annual demand at selling price $P_s, s = 0, 1$;
- PR_0 : retailer's maximum average annual profit when no discount is offered;
- PR_1 : retailer's average annual profit when a discount is offered and the EOQ (q_0^*) is used for orders placed when the discount is not in effect;
- q_0 : retailer's order quantity when no discount is offered;
- Q_1 : retailer's order quantity when a discount is offered;
- q_1 : portion of Q_1 sold at the non-discounted price of P_0 (equivalently, the *forward buy* quantity);
- t_0 : length of regular time interval between successive retailer setups;
- t_1 : length of time interval covered by retailer's special order of size Q_1;
- L: length of the planning cycle, i.e., time between supplier price promotions.

As noted previously, the retailer's demand rate depends on the price it offers consumers. We make three assumptions with respect to the retailer's demand function. First, the price elasticity of consumer demand under the normal price is different from that under a promotional (reduced) price. Bell and Lattin [20] and Blattberg, Briesch and Fox [24] provide empirical evidence in support of this assumption. The intuitive idea behind this assumption is that because a promotional discount is only available for a short time, consumers know that they have to respond quickly. Therefore, a higher degree of promotional price elasticity is expected as compared to the normal price elasticity. In contrast, under the normal price, consumers may anticipate that a price decrease will last for a considerable time, potentially leading to weaker price elasticity.

Under the normal scenario S_0, the price is stable over time, and the demand-price relationship is modeled using the following linear function:

$$D_0 = r_c - r_0 P_0 \ , \tag{2–1}$$
$$\text{with } r_c \geq 0, r_0 > 0 \ ,$$

where r_c is a constant and r_0 is the regular price elasticity of consumer demand. Under the scenario S_0, clearly it is optimal for the retailer to order its corresponding EOQ. Using Equation (2–1), the retailer's problem for maximizing profit under the normal scenario can be written as:

$$PR_0(P_0^*, q_0^*) = \max \left[(P_0 - c)D_0 - \frac{icq_0}{2} - \frac{AD_0}{q_0} \right] \tag{2–2}$$

subject to:

$$D_0 = r_c - r_0 P_0$$
$$\text{with } r_c \geq 0, r_0 > 0,$$

where P_0^* and q_0^* are the optimal price and the optimal order quantity, respectively. Arcelus and Srinivasan [12] showed that the optimal solutions to (2–2) can be written as:

$$q_0^* = \sqrt{\tfrac{2AD_0^*}{ic}} \tag{2–3}$$
$$P_0^* = \tfrac{1}{2}\left[\left(c + \tfrac{A}{q_0^*}\right) + \tfrac{r_c}{r_0}\right] \ ,$$

where D_0^* is the demand rate corresponding to the optimal retailer price. Thus, under the normal scenario S_0 without any supplier price discounts, the retailer replenishes its inventory every $t_0^* = q_0^*/D_0^*$ time units.

Second, we assume that the consumer population consists of two groups: brand *loyals* and *impulsive consumers*. The brand loyals always buy the product, whereas impulsive consumers do not. Moreover, impulsive consumers tend to make buying decisions without thoughtful and deliberate consideration which makes their buying

behavior more easily altered by the featured point-of-purchase promotions. We take impulsive consumers into consideration, because impulsive buying is a prevalent phenomenon in the retailing industry. Reports show between 27% and 62% of all department store purchases are identified as "unplanned" purchases [17], and 88% impulse purchases are because shoppers find products are offered at a lower price (or on sale) [45]. When a temporary price discount is offered, the impacts of a price discount on brand loyals can be decomposed into two aspects: purchase acceleration and forward buying. In other words, the demand rate of brand loyals is going to increase at the rate r_0 units per dollar of price reduction as in normal scenario S_0; moreover, these consumers are willing to carry some additional inventory in return for a reduction in price. In addition to the purchase acceleration and forward buying of brand loyals, impulsive consumers also temporarily buy the product at the discounted price for several reasons: some impulsive consumers may enjoy the process of looking for deals, some may want to try the product at a lower risk level, and some may simply be driven by a sudden and spontaneous desire or urge to buy things on sale.

Third, we assume that the relative frequency of price promotions impacts the performance of the promotions. For example, suppose the retailer provides a price promotion every L time units, and the promotion lasts l units of time. If $l \geq L$, then this is no longer a high-low pricing strategy, and it degenerates into an EDLP strategy, so that the price elasticity is equivalent to that under the normal price. On the other hand, if $l < L$, a smaller value of l implies a stronger price elasticity in response to promotions. We use this assumption to capture the phenomenon, shown by Blattberg and Neslin [26], that a temporary price reduction can substantially boost sales, but frequent usage of promotions diminishes their effectiveness.

We therefore model the short-term relationship between price and demand under the promotional scenario S_1 as follows:

$$D_1 = D_0^* + \delta(r_0 + \tilde{r}_f(l) + \tilde{r}_i(l)) \ , \tag{2–4}$$

where

$$0 < l \leq L, 0 \leq \delta = P_0^* - P_1 \ ,$$

and where D_1 measures the consumer demand per unit time under the promotional price; $\tilde{r}_f(l)$ and $\tilde{r}_i(l))$ denote, respectively, the brand loyals' forward buying and the impulsive consumers' number of units purchased per dollar of price reduction; r_0 is the brand loyals' purchase acceleration per dollar of price reduction (which is simply the regular price elasticity), and δ is the discount offered by the retailer (in dollars). For ease of notation, we will sometimes suppress the dependence of the functions $\tilde{r}_f$ and $\tilde{r}_i$ on l. We assume that $\tilde{r}_f$ is a decreasing function of the promotion duration l for a given planning cycle length L, which implies a lower degree of forward buying corresponding to a longer promotion duration. We also assume that $\tilde{r}_f$ equals zero when $l = L$, which implies that when the price is constant during the entire planning cycle length, forward buying will cease. In our computational tests in Section 2.3, we use the functional form $r_f\left(e^{\beta\left(1-\frac{l}{L}\right)} - 1\right)$, where r_f and β correspond to positive scaling parameters. The function $\tilde{r}_i$ is also expressed as a function of l, and is also decreasing in l for a given L, as the degree of impulsive buying decreases in the relative duration of the promotion period. We also assume that $\tilde{r}_i = 0$ when $l = L$. In our computational tests in Section 2.3, we use the functional form $r_i(e^{\gamma\left(1-\frac{l}{L}\right)} - 1)$, where r_i and γ correspond to positive scaling parameters. Observe that when $l = L$, Equation (2–4) is equivalent to (2–1).

When the retailer's promotion period expires (after l time units, assuming $l < L$), consumer demand temporarily falls to zero, as impulsive consumers stop purchasing the product and loyals start consuming their accumulated stockpiles as a result of forward buying. The consumer demand rate returns to normal when the brand loyals' inventories have run out (Figure 2-1).

2.2.2 Retailer Profit under Promotions

In this section we define a model for the retailer's average annual cost. The expression for the retailer's optimal profit under the scenario S_0, denoted as PR_0, was

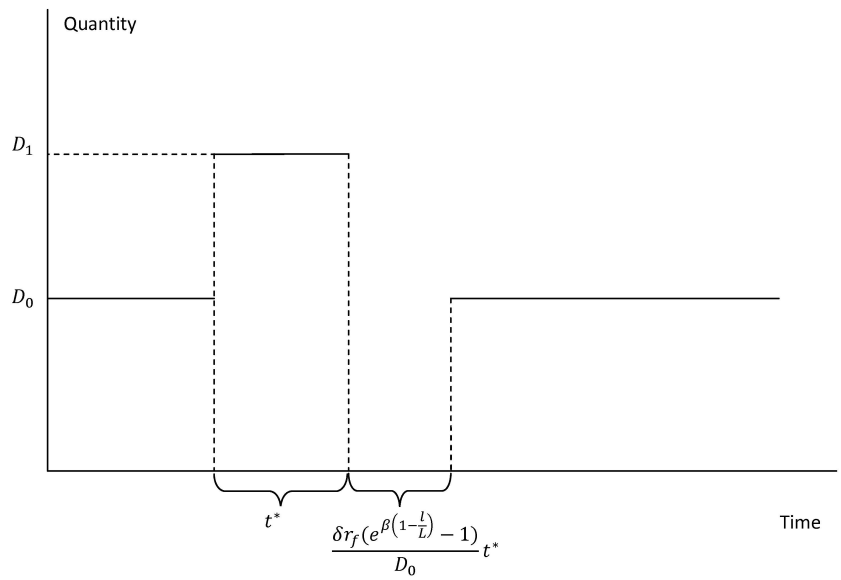

Figure 2-1. Consumer demand over time.

provided in Section 2.2.1, and the optimal value of PR_0 can be determined by finding the corresponding stationary point of the objective function. In the rest of this section, we focus on constructing a mathematical model for the retailer's average profit per unit time under the promotion scenario S_1, denoted as PR_1.

As noted previously, we assume that the supplier periodically offers a discount to the retailer. For model tractability, we assume that the time between promotions is long when compared to the retailer's normal replenishment cycle length t_0 and that the length of the planning cycle must be an integer multiple of t_0, i.e., $L = nt_0$ for some $n \in \mathbb{N}$, where the value of n is predetermined. That is, we assume an exogenously determined time between promotions (e.g., yearly or quarterly).

Under the scenario S_1, at the beginning of the planning cycle, the supplier offers a price cut, d, to the retailer, and the retailer then requests a large order of size Q_1, which will cover demand over a time interval, t_1, which is at least as long as the regular replenishment time interval, t_0. Furthermore, the length of this special time interval is restricted to an integer multiple of of t_0, i.e., $t_1 = mt_0; m \in \mathbb{N}$, where m is a decision

variable in our model. This assumption is made strictly in order to facilitate computation of the bullwhip effect as we later discuss in Section 2.2.3. Clearly m should be smaller than n (otherwise the system would not ever return to the normal operating scenario). We assume that the retailer will sell $Q_1 - q_1$ of the units purchased at a discounted price to its consumers. That is, after purchasing Q_1 units at a discount from its supplier, the retailer then discounts its price to its consumers for $Q_1 - q_1$ of these units. Clearly the case of $q_1 = Q_1$ implies that the retailer does not pass any of the discount on to its consumers. Note that q_1 thus determines the amount of the retailer's *forward buy*.

The retailer's profit per promotion cycle under the scenario S_1 is composed of three parts (as shown in Figure 2-2). The first component considers the time during which the retailer offers a discount to its consumers, which has a duration of $l = (Q_1 - q_1)/D_1$ time units. Included here are:

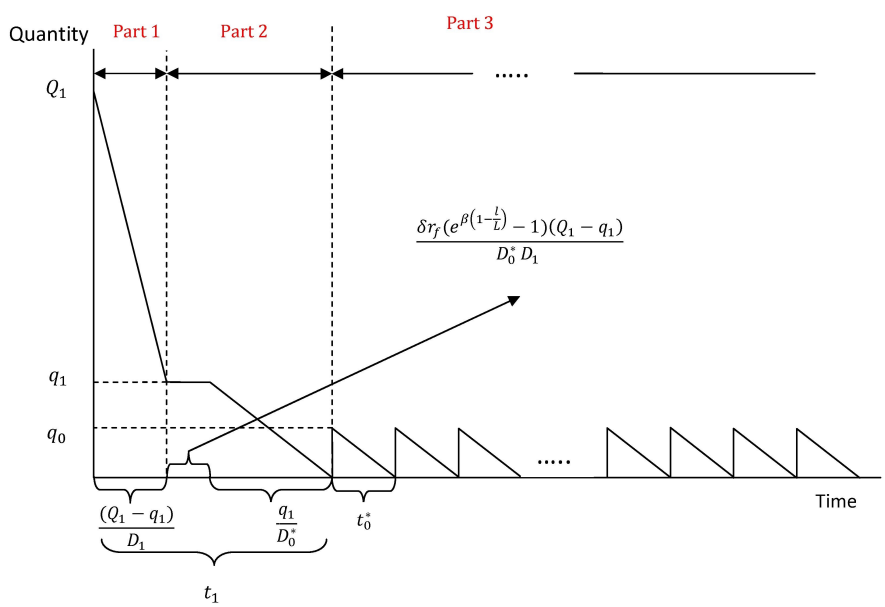

Figure 2-2. Retailer's inventory.

- the revenue from selling the item at the discounted price P_1;
- the cost of purchasing $(Q_1 - q_1)$ units of the item at the discounted cost $(c - d)$ per unit;
- the setup cost for placing the order, A; and
- the cost of holding an average of $(Q_1 + q_1)/2$ units over a period of length $(Q_1 - q_1)/D_1$ at a rate of $i(c - d)$ per unit.

The second part, of length $\delta\tilde{r}_f(Q_1 - q_1)/(D_0^* D_1) + q_1/D_0^*$, accounts for the profit associated with any non-discounted part of the retailer's special order. That is, during this time, the retailer has eliminated its price discount, but has remaining inventory from its special purchase of size Q_1. Included here are:

- the revenue from selling q_1 units at a non-discounted price P_0^*;
- the cost of purchasing q_1 units at the discounted cost of $(c - d)$ per unit;
- the holding cost for q_1 units for a period of length $\delta\tilde{r}_f(Q_1 - q_1)/(D_0^* D_1)$ at a rate of $i(c - d)$ per unit; and
- the holding cost for an average of $q_1/2$ units over a period of length q_1/D_0^* at a rate of $i(c - d)$ per unit.

The third and final part, of length $(nt_0^* - (Q_1 - q_1)/D_1 - \delta\tilde{r}_f(Q_1 - q_1)/(D_0^* D_1) - q_1/D_0^*)$ covers the profit resulting from selling at the normal price P_0^* while ordering in lots of size q_0^* for the remainder of the planning horizon. We employ the time proportional cost assumption of Naddor [76], i.e., we assume that the profit associated with this third component is proportional to the time remaining in the promotion cycle and accumulates at an annual rate given by PR_0 (Equation (2–2)).

Combining all of the above yields the generalized model below:

$$PR_1(Q_1^*, q_1^*, \delta^*, m^*) = \max\left\{\frac{1}{L}\left\{(P_0^* - \delta)(Q_1 - q_1) + P_0^* q_1 - (c - d)Q_1 - A\right.\right.$$
$$\left.\left. - \left[\frac{Q_1^2 - q_1^2}{D_1} + \frac{q_1(Q_1 - q_1)\delta\tilde{r}_f}{D_0^* D_1} + \frac{q_1^2}{D_0^*}\right] \times \frac{i(c - d)}{2}\right\} + \left(1 - \frac{m}{n}\right) PR_0\right\} \tag{2–5}$$

subject to:

$$\frac{Q_1 - q_1}{D_1} + \frac{\delta \tilde{r}_f (Q_1 - q_1)}{D_0^* D_1} + \frac{q_1}{D_0^*} = m \frac{q_0^*}{D_0^*} \tag{2–6}$$

$$D_1 = D_0^* + \delta(r_0 + \tilde{r}_f + \tilde{r}_i) \tag{2–7}$$

$$l = \frac{Q_1 - q_1}{D_1} \tag{2–8}$$

$$L = \frac{n q_0^*}{D_0^*} \tag{2–9}$$

$$0 \leq q_1 \leq Q_1, \delta \geq 0, m \leq n, m \in \mathbb{N} \ . \tag{2–10}$$

The above optimization problem is a mixed integer nonlinear programming problem without a closed-form solution. Even if we relax the integrality constraint on m, because the objective function is neither convex or concave, we can only use the necessary KKT conditions to yield candidate KKT points for an optimal solution. Thus, the retailer's problem as we have formulated it is a nonconvex mixed integer optimization problem, and we cannot, therefore, guarantee finding an optimal solution.

In practice, however, it is typically not necessary for the retailer to solve this nonconvex mixed integer optimization problem exactly for two reasons. First, a retailer rarely offers, say, an 18.79% discount, for example, because consumers are generally insensitive to very small changes in price; it is not uncommon for a retailer to round the "optimal" discount to the nearest multiple of 5% or 1%. Second, the length of the promotion period is often stated as a multiple of some base planning period length, e.g., one day or one week. Consequently, the retailer only needs to consider a finite number of discrete combinations of l and δ, the promotion duration and the discount level, for any given value of the manufacturer's discount d. Thus, it is not impractical for the retailer to enumerate candidate values of l and δ in order to solve the resulting optimization problem.

After fixing l and δ, we can show that under very mild conditions on $\tilde{r}_f$ and $\tilde{r}_i$, the continuous relaxation of the retailer's problem is a concave program (Appendix A for proof of the concavity of the objective, assuming $\tilde{r}_i + r_0 \geq \tilde{r}_f$ and $\delta \tilde{r}_f \leq 4D_0$).

Moreover, this optimization problem contains only a single integer variable, m. Thus, for given values of l and δ, we can easily solve the retailer's mixed integer optimization problem by first solving the relaxation, then rounding the resulting value of m up and down, and solving the convex program that results at each of these fixed values of m (the subproblem with the higher objective function value provides the retailer's optimal solution). That is, we first solve the relaxed convex program and, if m^* is integer, we stop with an optimal solution. Otherwise, we solve two additional convex programs with $m = \lfloor m^* \rfloor$ and $m = \lceil m^* \rceil$, respectively, and select the one with the higher objective function value. The Appendix B demonstrates that this approach provides an optimal solution for fixed l and δ values.

2.2.3 Supplier Profit

This section discusses the costs incurred and revenues received by the supplier in the two-stage system we have described. We assume the supplier incurs transportation costs in ensuring delivery of items to the retailer, and that the supplier also stocks inventory (and incurs associated holding costs) in order to meet retailer demand. The following subsection discusses the structure of supplier transportation-related costs. Following this, we consider the supplier's inventory-related costs and the problem of setting a wholesale discount level associated with a promotion. These factors combine to determine the supplier's overall profit function.

2.2.3.1 Supplier transportation cost

The bullwhip effect hampers a supplier's ability to efficiently utilize capacity. Our model therefore considers how this affects both a supplier's inventory costs and its transportation capacity. Although we explicitly discuss transportation capacity, this capacity might correspond to any type of generic capacity, including production capacity. To compute the transportation cost (which we assume is borne by the supplier), we use the following notation:

- $W =$ regular in-house truck capacity in units;

- $SF_R(W) =$ fixed cost per shipment as a function of regular capacity;
- $SF_0 =$ fixed cost per outside or emergency shipment;
- $SP_R =$ regular unit shipping cost;
- $SP_0 =$ outside carrier unit shipping cost;
- $T_Q(W) =$ per shipment cost if the size of delivery is W;
- $T(W) =$ total transportation cost per promotion cycle.

Ernst and Pyke [46] explored optimal inventory and shipping policies for a two-level distribution system composed of a warehouse and a retailer with random consumer demand. We use a similar transportation cost structure in our model, which considers a supplier's in-house trucking cost as well as costs associated with an outside source (common carrier). In-house trucking corresponds to a fleet operated by the supplier. We assume that the supplier also contracts with an outside trucking firm so that if the in-house trucking capacity is not sufficient to deliver the retailer's order, the supplier must use this outside trucking capacity, at a higher cost, to ship the difference (alternatively, the supplier may arrange for an unplanned shipment, which typically comes at an additional cost). This implies that we assume that the supplier must meet all of the retailer's orders without backordering any deliveries.

Shipping costs contain several components. First, we assume a fixed cost per shipment using in-house capacity, and that this fixed cost is a linear function of in-house truck capacity, which implies that in-house truck capacity is one of supplier's decision variables. Second, we consider a cost per unit shipped using in-house truck capacity, and assume that this cost is independent of the fixed cost. Moreover, we assume that the outside trucking firm charges both a fixed cost per shipment and a cost per unit shipped. Thus, the supplier's per-shipment transportation cost is

$$\begin{aligned} T_Q(W) &= SF_R(W) + SP_R \times \min(Q, W) + SF_O \times I(Q - W > 0) \\ &\quad + SP_O \times \min(Q - W, 0) \ , \end{aligned} \tag{2–11}$$

where Q is the retailer's order quantity, and $I(Q-W>0)$ is the identity function, defined as

$$I(Q-W>0)=\begin{cases}0, & Q-W\leq 0,\\ 1, & Q-W>0\ .\end{cases}$$

As noted before, we approximate the fixed cost of in-house shipping using a linear function

$$SF_R(W)=aW, \tag{2–12}$$

where a is a scalar that denotes the fixed cost per unit of truck capacity, charged on a per-shipment basis.

Since the system demand is deterministic, after the retailer's inventory policy has been set, the supplier can evaluate the optimal in-house trucking capacity independently of its internal inventory replenishment decisions (because the supplier's inventory cost is independent of its outbound transportation capacity, which depends only on the retailer's order quantities).

For the normal scenario S_0, it is straightforward for the supplier to minimize transportation cost. Since there is only a single order size equal to q_0^*, the supplier's in-house trucking capacity can be either 0 or q_0^*, and a rational supplier chooses the one that results in lower cost. For scenario S_1, there are two kinds of deliveries per promotion cycle: a large size delivery, Q_1 and $(n-m)$ regular deliveries of size q_0^*. We therefore need solve the following optimization problem to determine the optimal in-house capacity:

$$T(W^*)=\min\{T_{Q_1^*}(W)+(n-m)t_0^*T_{q_0^*}(W)\}\ , \tag{2–13}$$

where $T_{Q_1^*}(W)$ and $T_{q_0^*}(W)$ are the supplier's per shipment costs if the delivery sizes equal Q_1^* and q_0^*, respectively, and $T(W)$ is the total transportation cost per promotion. The above objective function is a piecewise linear function with breakpoints at q_0^* and Q_1^* with $\lim_{W\to\infty}T(W)=\infty$; thus, an optimal solution can only occur at one of the three

critical points 0, q_0^*, and Q_1^*, and we need only compare the values of the function at these points and choose the minimum.

2.2.3.2 Supplier inventory cost and profit

The notation we use to construct the supplier's inventory cost model is summarized as follows:

- $A_S =$ supplier's setup cost;
- $i_S =$ supplier's inventory holding cost rate per unit per year;
- $T_0 =$ length of supplier's regular replenishment period;
- $T_1 =$ length of supplier's special replenishment period;
- $c_S =$ supplier's regular production cost;
- $d_S =$ discount on raw material;
- $K_0 = T_0/t_0$;
- $K_1 = (T_1 - t_1)/t_0 + 1$;
- $PS_0 =$ supplier's optimal yearly profit when no discount is offered;
- $PS_1 =$ supplier's profit when discount is offered and an optimal policy for PS_0 is used.

In addition to transportation cost, we also consider inventory-related costs in computing the supplier's profit. Under the scenario S_0, we assume that the supplier uses a discrete EOQ quantity denoted by Q_S. It is easy to show that Q_S may be constrained to an integer multiple of q_0^* (Goyal [53], who used a similar assumption to study a two-stage inventory model). We can determine the supplier's optimal profit under the normal scenario by solving the following maximization problem:

$$PS_0(K_0^*) = \max\left\{(c - c_S)D_0^* - \left[\frac{A_S}{K_0 t_0^*} + \frac{(K_0 - 1)D_0^* t_0^* i_S c_S}{2}\right] - \frac{T_{q_0^*}(W^*)}{t_0^*}\right\} , \qquad (2\text{–}14)$$

where K_0 is the supplier's multiple of the retailer's normal order quantity in a normal production run, and $T_{q_0^*}(W^*)$ is the minimum per shipment transportation cost when the

retailer's order size is q_0^*, which is determined independently by the retailer. Goyal [53] showed that the supplier's optimal production multiple K_0^* satisfies:

$$K_0^* \times (K_0^* + 1) \geq \frac{2A_S}{D_0^* i_S (t_0^*)^2} \geq K_0^* \times (K_0^* - 1) \ . \tag{2–15}$$

We extend this to the case in which the retailer places a special order of size Q_1^*, which implies that the supplier needs to place a special order with a quantity equal to one of the values Q_1^*, $Q_1^* + q_0^*$,$Q_1 + 2q_0^*$, ... After exhausting this special production batch, the supplier's production is assumed to return to its normal size, $K_0^* q_0^*$, which is the optimal solution derived from (2–15).

In addition to transportation cost, the supplier's profit per promotion cycle contains two parts (as shown in Figure 2-3). The first part, of length T_1, consists of the profit associated with the special order at a discounted raw material price. Included here are:

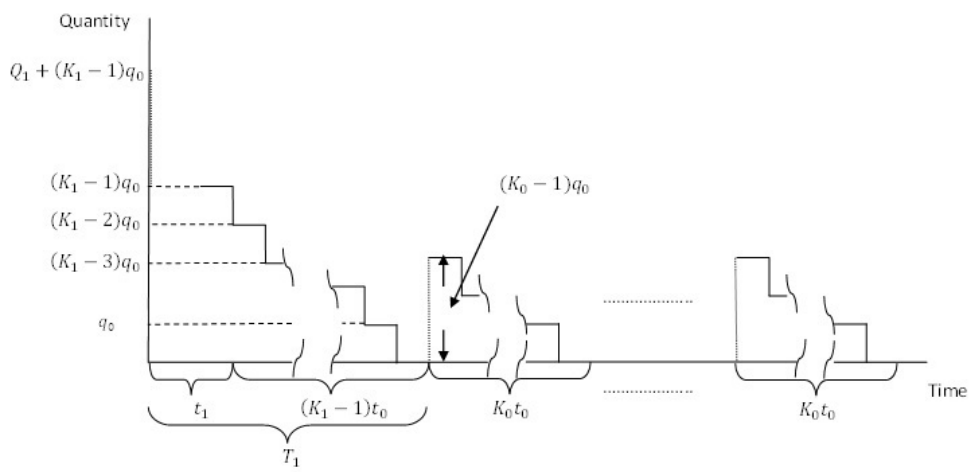

Figure 2-3. Supplier's inventory.

- the revenue from selling the special order of size Q_1^* at $(c - d)$ per unit and $(K_1 - 1)q_0^*$ units at c per unit;
- the cost of producing $Q_1^* + (K_1 - 1)q_0^*$ units at a cost of $(c_S - d_S)$;
- the fixed setup cost, A_S;

- the cost of holding $(K_1 - 1)q_0^*$ units for a period of length t_1^* at a rate $(c_S - d_S)i_S$; and
- if $K_1 \geq 2$, the cost of holding an average of $(K_1 - 2)q_0^*/2$ units for a period of $(K_1 - 1) \times t_0$ at a rate $(c_S - d_S)i_S$ per unit.

The second part, of length $(nt_0^* - T_1)$, covers the profit arising from the return to regular production for the remainder of the promotion cycle at a rate of PS_0 per year (Equation (2–14)).

Combining each of the above components yields the supplier's profit-maximizing model below:

$$\begin{aligned} PS_1(K_1^*) = \ & \tfrac{1}{nt_0^*}\max\{(c-d)Q_1 + (K_1-1)q_0^*c - (c_S - d_S)[Q_1 + (K_1-1)q_0^*] - A_S \\ & - \left[(K_1-1)t_1 + \tfrac{(K_1-1)(K_1-2)t_0^*}{2}\right] q_0^* i_S(c_S - d_S) \\ & + [nt_0^* - (t_1 + (K_1-1)t_0^*)]\left(PS_0(K_0^*) + \tfrac{T_{q_0^*}(W^*)}{t_0^*}\right) - T(W^\star)\Big\} \end{aligned} \tag{2–16}$$

subject to: $K_1 \in \mathbb{Z}$.

The above objective is concave in K_1, and so we can easily show that the optimal special production size K_1^* satisfies the following inequalities

$$\max\left\{1, \tfrac{q_0^*(c-c_S+d_S)-t_0^*PS_0}{q_0^* i_S(c_S-d_S)t_0^*} - \tfrac{t_1}{t_0^*} + 1\right\} \leq K_1^* \leq \max\left\{1, \tfrac{q_0^*(c-c_S+d_S)-t_0^*PS_0}{q_0^* i_S(c_S-d_S)t_0^*} - \tfrac{t_1}{t_0^*} + 2\right\}$$
$$\text{such that } K_1 \in \mathbb{Z} \ . \tag{2–17}$$

2.2.4 Quantifying the Bullwhip Effect

In this study, we use the ratio of standard deviation of order quantity to standard deviation of demand $\left(\frac{std[Q]}{std[D]}\right) \equiv BWE$, as a measure of the bullwhip effect, where Q is the retailer's order quantity at the beginning of each regular replenishment period of length t_0 and D is the consumer demand that occurs during this same interval length. (Because the average consumer demand per unit time must equal the average retailer-to-supplier order quantity per unit time in the long run, this is equivalent to considering the ratio of coefficient of variation values at the stages in our model.)

Under the normal scenario S_0, the retailer repeatedly orders q_0 at the beginning of each regular replenishment period, and so under this scenario, the retailer's order quantity has zero variance. Moreover, the consumer demand occurs at a constant rate, so the consumer demand rate also has zero variance. Consequently, the bullwhip effect is non-existent under the normal scenario S_0.

Under the promotion scenario S_1, when presented with the supplier's price discount, the retailer places a special order of size Q_1 at the beginning of the promotion cycle, which covers mt_0 regular replenishment periods, which implies that the retailer places no order at the beginning of the following $m-1$ regular replenishment periods. After depleting the special order quantity Q_1, the retailer returns to the normal strategy and orders q_0 at the beginning of the remaining $n-m$ replenishment regular replenishment periods. Combining these observations, the mean and standard deviation of the retailer's order quantity can be calculated as:

$$E[Q] = \frac{q_0(n-m) + Q_1}{n} \quad , \tag{2–18}$$

$$std[Q] = \sqrt{\frac{(q_0 - E[Q])^2(n-m) + (Q_1 - E[Q])^2 + (0 - E[Q])^2(m-1)}{n}} \quad . \tag{2–19}$$

Under the promotion scenario S_1, the consumer demand is composed of three parts (as shown in Figure 2-1). The first component has a duration of $t^* = (Q_1 - q_1)/D_1$ time units and a demand rate of D_1; the second component has a duration of $t^{**} = \delta\tilde{r}_f(Q_1 - q_1)/(D_0 D_1)$ time units and a demand rate of zero; and the third component has a duration of $t^{***} = q_1/D_0 + (n-m)t_0$ time units and a demand rate of D_0. Because t^*, t^{**}, and t^{***} may not be integer multiples of t_0, we need to define additional parameters before computing the demand variance. To do this, we define the following quantities:

- $n_1 = \left\lfloor \frac{t^*}{t_0} \right\rfloor$;
- $n_2 = \max\{\left\lfloor \frac{t^{**}}{t_0} + \frac{t^*}{t_0} - n_1 - 1 \right\rfloor, 0\}$;

- $n_3 = \left\lfloor \frac{t^{***}}{t_0} \right\rfloor$;
- $l_1 = \begin{cases} 1 & \frac{t^{**}}{t_0} + \frac{t^{*}}{t_0} - n_1 < 1, \\ 0 & \text{otherwise;} \end{cases}$
- $l_2 = \left\lceil \frac{t^{***}}{t_0} - n_3 \right\rceil$.

A promotion cycle begins with n_1 regular replenishment intervals, each with demand $D_1 t_0$, and ends with n_3 regular replenishment periods, each with demand $D_0 t_0$. For the rest of the promotion cycle, there are two possible cases: (i) $l_1 = 0$, which implies the duration of time with a zero demand rate is less than the regular replenishment interval length t_0, and a single interval exists that either begins with a demand rate of D_1 and ends at zero, begins with a zero demand rate and ends with a demand rate of D_0, or begins with a demand rate of D_1 and ends with a demand rate of D_0. The demand during this interval equals $d_3 = (\frac{t^*}{t_0} - n_1)D_1 + (1 + n_1 - \frac{t^*}{t_0} - \frac{t^{**}}{t_0})D_0$; (ii) $l_1 = 1$, which implies that there are $n - n_1 - n_3$ periods of length t_0 containing intervals of zero demand. Included here are:

- a regular replenishment period starting with a demand rate D_1 and ending with a zero demand rate, which has total demand of $d_4 = (\frac{t^*}{t_0} - n_1)D_1$;
- n_2 regular replenishment periods containing zero demand;
- a regular replenishment period starting with a demand rate 0 and ending with a demand rate D_0, which has a total demand $d_5 = \left(\frac{t^{***}}{t_0} - n_3\right) D_0$.

Combining each of the above components, the mean and variance of consumer demand are computed as

$$E[D] = \frac{n_1 D_1 t_0 + n_3 D_0 t_0 + (1 - l_1) d_3 + l_1 d_4 + l_2 l_1 d_5}{n}, \tag{2–20}$$

$$\begin{aligned} Var[D] = \quad & \tfrac{1}{n}\{n_1(D_1 t_0 - E[D])^2 + n_2(0 - E[D])^2 + n_3(D_0 t_0 - E[D])^2 \\ & +(1 - l_1)(d_3 - E[D])^2 + l_1(d_4 - E[D])^2 + l_2 l_1 (d_5 - E[D])^2\}, \end{aligned} \tag{2–21}$$

and we have $std[D] = \sqrt{Var[D]}$.

In summary, our model considers two scenarios, a normal scenario and a promotion scenario, and we assume that the retailer and supplier optimize their profits independently. We first compute the retailer's optimal profit, PR_0, and then the supplier's optimal profit, PS_0, under the normal scenario. For the promotion scenario, given a temporary supplier price discount, the retailer optimizes its profit, PR_1, by choosing appropriate values of Q_1, q_1 and δ. Given the retailer's ordering policy, the supplier chooses its in-house trucking capacity W and arranges a special production order (determined by K_1) to maximize its profit, PS_1. Finally, we compute the bullwhip effect, $BWE = \frac{std[Q]}{std[D]}$, the retailer's profit gain, $\Delta_R = \frac{PR_1 - PR_0}{PR_0} \times 100\%$, the supplier's profit gain, $\Delta_S = \frac{PS_1 - PS_0}{PS_0} \times 100\%$, and the total system profit gain, $\Delta = \frac{PS_1 + PR_1 - PS_0 - PR_0}{PS_0 + PR_0} \times 100\%$.

2.3 Numerical Experiments

The goal of our numerical experiments in this chapter is to characterize how price discounts influence the system profit. We are particularly interested in the conditions under which increased system profits can coexist with the bullwhip effect. To gain some insight into this question, we performed numerous computational tests that parameterize on the discount given by the supplier, d.

In order to cover a broad range of potential product characteristics, we consider three categories of inventory based on unit values. We apply an $80/20$ Pareto principle and classify items based on annual dollar movement (as measured by cD) as follows:

- Category A items comprise 20% of the SKUs and contribute to 80% of annual dollar movement.
- Category B items comprise 30% of the SKUs and contribute to 15% of annual dollar movement.
- Category C items comprise 50% of the SKUs and contribute to 5% of annual dollar movement.

The retailer's standard parameter values for Category B items (except for β, γ and r_f) correspond to those used by Ardalan [13]. The values of β, γ and r_f for

Category B items were chosen based on the work of Gupta [54], which estimates that only 16% of the sales increase from promotions results from brand loyals' forward buying. We note that it is possible to "stack the deck" in favor of promotions by creating numerical examples with an extremely high rate of demand from impulsive consumers (by choosing a high value of γ). As we later discuss, however, the parameters we have chosen are reasonably conservative in terms of the impact of impulsive consumers on the profit from promotions. According to the study by Blattberg and Neslin [26], the increase in sales due to price cuts can be as much as 10 times the normal level (or a 1000% increase). However, the parameters we have chosen lead to problem instances in which the demand from impulsive consumers does not exceed 21% of total demand. Thus we can create instances in which the profit from promotions is much higher by increasing the demand from impulsive consumers (although we have chosen not to do this). The supplier's standard parameter values for category B items were chosen to reflect the fact that the supplier usually has lower holding costs and higher setup costs than the retailer. As shown in Table 2-1, all product categories use the same values of $i = 0.25$, $i_s = 0.15$, $\beta = \ln(2)$ and $\gamma = \ln(1.33)$.

Based on the parameter values of Category B items, those for Category A and C items were computed according to the $80/20$ principle (as shown in Table 2-1). For example, we set the parameters for Category A items to ensure that, under the normal scenario, the annual demand for Category A items, D_A, equals two thirds of D_B. We then set the per unit cost of Category A items using $c_A = \frac{c_B \times D_B}{D_A} \times \frac{80}{15}$. The other cost-related parameters for Category A items were then computed by scaling the Category B parameters in proportion to the relative values of D_A and c_A. Finally, we set the regular price elasticity of consumer demand for Category A items using $r_{0_A} = \frac{r_{0_B} \times D_A}{D_B} \times \frac{15}{80}$; the value of the only remaining parameter, r_c, can be found by trial and error to ensure that the $80/20$ principle is satisfied. For each item category, we solved the model for various values of supplier discount percentage, i.e., $\frac{d}{c} \times 100\%$, ranging from 0.01 to

Table 2-1. Parameter settings for computational tests.

Category	c	c_S	i	i_S	A	A_S	d_S	α
A	80	48	0.25	0.15	80	480	4.8	4
B	10	6	0.25	0.15	10	60	0.6	0.5
C	2	1.2	0.25	0.15	2	12	0.12	1

Category	SP_R	SF_0	SP_0	β	γ	r_c	r_0	r_f	r_i
A	8	1200	12	$\ln(2)$	$\ln(1.33)$	42800	375	133	375
B	1	150	1.5	$\ln(2)$	$\ln(1.33)$	49000	3000	1000	3000
C	0.2	30	0.3	$\ln(2)$	$\ln(1.33)$	61600	15000	6700	15000

0.3. We apply the approach discussed in Section 2.2.2 to solve the retailer's problem for a discrete set of candidate values of δ and l under the promotion scenario, S_1. We examine thirty possible values of δ (from $c \times 1\%$ to $c \times 30\%$) and one hundred potential values of l (from 1 to 100). Tables 2-2 - 2-4 provide detailed results of the models for Categories A, B and C, and for given values of the supplier's discount d.

The results in the tables suggest the following. For all inventory categories, as the supplier discount, d, increases, the retailer's discount, δ, also increases, and, somewhat surprisingly, the retailer's optimal discount, δ, is not always smaller than the supplier's discount, d. Note that for all inventory categories, the retailer's special order size, Q_1, and corresponding profit gain, Δ_R, both increase as d increases, which indicates that when a sufficient number of impulsive consumers exist (determined by $\tilde{r}_i$), our results for the retailer are consistent with intuition and with the results of Arcelus and Srinivasan [12]. Thus, the retailer's incentive to press the supplier for periodic discounts is clear. Moreover, from a system perspective, a positive level of supplier discount is desirable for all item categories, which implies that the bullwhip effect does not always imply a system loss and can coexist with higher profit levels as compared with the case of EDLP. However, when the supplier and retailer operate independently, it will likely be the case that the supplier's optimal action is not optimal for the system, i.e., the discount offered by the supplier (if any) is less than the level of supplier discount that maximizes the system profit. In addition, it is possible that the supplier and the total system may be

worse off as a result of providing a discount to the retailer. Both the supplier's profit gain, Δ_S, and the total system profit gain, Δ, first increase in d, and then, after reaching a peak at some positive value of d, decrease in d. For Category C items, Δ_S is always negative, which indicates that the supplier should not discount C items under our model assumptions. Thus, the supplier's choice of whether or not to offer a discount depends on the item category. The last column of Tables 2-2 - 2-4 provides the percentage of impulsive consumers' demand to illustrate that our examples do not artificially inflate the extra demand from impulsive consumers when compared to practical cases.

We also observe that for all inventory categories, the ratio of the standard deviation of retailer orders to that of consumer demand, the BWE, is always greater than one, which indicates that bullwhip effect exists for our problem instances. Note that for all categories, the BWE tends to *increase* as the discount d increases, which is consistent with the findings of Lee, Padmanabhan, and Whang [64, 65] that price variations serve as one of the causes of bullwhip effect.

Thus, for the model we have developed, it is not only in the retailer's best interest to seek periodic discounts, but it is also often in the supplier's best interest to provide them, despite the associated increase in costs due to the bullwhip effect. We note, however, that the optimal independent decision for the supplier does not lead to the optimal system profit, i.e., the supplier and retailer can typically reap higher gains through coordinating their actions (although we leave the specifics of such coordination mechanisms for further research).

Table 2-5 summarizes average, minimum, and maximum profit changes and the BWE for each category. The numbers in parentheses in Table 2-5 represent the value of the supplier's discount at which the corresponding maximum (or minimum) profit increase for the supplier occurs. Although there appears to be no direct relationship between profit changes (Δ_R, Δ_S and Δ) and the BWE, item categories with generally smaller values of the BWE tend to produce higher gains as a result of a supplier

Table 2-2. Computational results for category A with standard parameter settings.

Category A								
$\frac{d}{c} \times 100\%$	$\frac{\delta}{c} \times 100\%$	Q_1	q_1	Δ_R	Δ_S	Δ	BWE	$SW\%$
1%	2%	731.68	210.38	0.10%	0.18%	0.15%	11.49	0.25%
2%	3%	1026.91	192.46	0.28%	0.37%	0.33%	9.22	0.56%
3%	3%	1293.7	182.2	0.53%	0.32%	0.42%	10.29	0.75%
4%	3%	1560.26	172.23	0.84%	0.19%	0.49%	11.23	0.93%
5%	3%	1826.58	162.57	1.23%	−0.03%	0.54%	12.08	1.11%
6%	4%	2469.39	117.14	1.72%	0.19%	0.89%	10.68	1.97%
7%	4%	2748.5	105.42	2.29%	−0.19%	0.93%	11.22	2.21%
8%	4%	3027.31	94.09	2.94%	−0.67%	0.97%	11.74	2.45%
9%	4%	3583.99	72.55	3.68%	−1.26%	0.99%	12.70	2.92%
10%	5%	3973.53	259.51	4.53%	−1.53%	1.23%	11.30	3.65%
11%	5%	4264.6	246.75	5.48%	−2.29%	1.24%	11.66	3.95%
12%	5%	4845.61	222.58	6.54%	−3.22%	1.22%	12.41	4.53%
13%	6%	5499.6	627.25	7.70%	−4.07%	1.28%	12.39	5.43%
14%	6%	5802.5	613.56	9.00%	−5.17%	1.28%	12.82	5.78%
15%	6%	6330.04	825.5	10.41%	−6.72%	1.07%	13.74	6.12%
16%	7%	6784.8	1000.6	11.95%	−7.63%	1.28%	12.88	7.14%
17%	7%	7413.9	971.73	13.64%	−9.23%	1.18%	13.52	7.94%
18%	7%	7952.76	1183.3	15.46%	−11.23%	0.92%	14.25	8.33%
19%	8%	8670.84	1577.44	17.42%	−13.04%	0.82%	13.65	9.52%
20%	8%	9096.84	1323.46	19.55%	−14.71%	0.88%	14.00	10.41%
21%	8%	9646.59	1535.11	21.83%	−17.29%	0.51%	14.70	10.86%
22%	9%	10402.23	1923.95	24.28%	−19.73%	0.30%	14.36	12.21%
23%	9%	11074.47	1895.92	26.91%	−22.56%	−0.05%	14.95	13.20%
24%	9%	11634.7	2108.09	29.71%	−25.77%	−0.52%	15.55	13.69%
25%	10%	12548.79	2253.73	32.71%	−28.80%	−0.81%	14.99	15.75%
26%	10%	13119.81	2466.05	35.92%	−32.47%	−1.35%	15.57	16.29%
27%	10%	14034.6	2667.92	39.31%	−37.06%	−2.31%	16.48	17.35%
28%	11%	14875.32	3047.26	42.94%	−41.05%	−2.83%	16.09	19.09%
29%	11%	15456.07	3260.84	46.79%	−45.44%	−3.47%	16.63	19.67%
30%	11%	16165.03	3240.31	50.86%	−50.01%	−4.11%	17.20	20.82%

discount (for both the retailer and supplier), which implies that actions other than stabilizing price should perhaps be taken to mitigate the negative impacts of the bullwhip effect.

Using our base parameter settings, we next examine how changes in specific parameter values affect the numerical results. For a parameter of interest, e.g., the retailer's setup cost A, we fix the values of all other parameters and consider how changing the value of the parameter of interest affects the results. That is, for the setup cost A, we consider the values A_H and A_L, where A_H is higher than the standard value of A, and A_L is lower than the standard value of A; we then compute the corresponding

Table 2-3. Computational results for category B with standard parameter settings.

Category B								
$\frac{d}{c} \times 100\%$	$\frac{\delta}{c} \times 100\%$	Q_1	q_1	Δ_R	Δ_S	Δ	BWE	$SW\%$
1%	3%	893.94	257.91	0.09%	−0.01%	0.05%	11.62	0.25%
2%	4%	1558.13	230.87	0.22%	0.25%	0.24%	11.57	0.67%
3%	4%	1878.28	220.65	0.42%	0.14%	0.29%	12.64	0.83%
4%	4%	2198.18	210.73	0.67%	−0.06%	0.34%	13.61	1.00%
5%	4%	2837.2	191.81	0.99%	−0.30%	0.41%	15.34	1.32%
6%	5%	3258.18	159.77	1.36%	−0.34%	0.60%	13.32	1.85%
7%	5%	3917.59	138.32	1.80%	−0.79%	0.64%	14.54	2.25%
8%	5%	4575.74	118.32	2.31%	−1.40%	0.65%	15.67	2.65%
9%	5%	4904.35	108.85	2.88%	−2.12%	0.65%	16.21	2.84%
10%	5%	5560.62	90.95	3.53%	−3.01%	0.61%	17.27	3.23%
11%	6%	6287.9	603.93	4.25%	−3.85%	0.64%	16.33	3.87%
12%	6%	6689.11	309.2	5.06%	−4.71%	0.70%	16.44	4.33%
13%	6%	7301.49	574.77	5.95%	−6.17%	0.54%	17.53	4.56%
14%	7%	8060.25	1082.9	6.93%	−7.41%	0.53%	16.97	5.32%
15%	7%	8754.83	1062.53	7.99%	−8.98%	0.42%	18.05	5.85%
16%	7%	9376.37	1327.92	9.15%	−10.87%	0.21%	19.14	6.11%
17%	8%	10166.07	1831.78	10.41%	−12.56%	0.16%	18.73	6.98%
18%	8%	10878.43	1811.6	11.77%	−14.61%	0.00%	19.59	7.57%
19%	8%	11508.78	2077.15	13.23%	−17.00%	−0.27%	20.52	7.87%
20%	9%	12419	2292.49	14.81%	−19.14%	−0.34%	19.71	9.18%
21%	9%	13333.24	2832.87	16.50%	−22.49%	−0.90%	20.96	9.50%
22%	9%	14060.58	2815.71	18.31%	−25.39%	−1.20%	21.69	10.15%
23%	10%	14920.38	3311.83	20.25%	−28.35%	−1.45%	20.78	11.28%
24%	10%	15940	3570.03	22.31%	−32.36%	−2.09%	21.86	11.99%
25%	11%	16819.29	4064.47	24.50%	−35.82%	−2.42%	21.47	13.19%
26%	11%	17855.86	4323.56	26.85%	−40.38%	−3.16%	22.60	13.96%
27%	11%	18510.64	4591.49	29.33%	−44.51%	−3.63%	23.34	14.34%
28%	12%	19693.62	5358.75	31.97%	−49.55%	−4.42%	23.05	15.64%
29%	12%	20745.33	5620.47	34.77%	−54.98%	−5.30%	24.10	16.47%
30%	12%	21794.31	5884.83	37.72%	−60.71%	−6.22%	25.18	17.28%

values of $\Delta_R, \Delta_S, \Delta$, and BWE. Table 2-6 shows the high and low parameter settings for each category of items. Figures 2-5 – 2-8 illustrate the results of these experiments, in which X_H (X_L) corresponds to the results from an experiment with parameter X at a high (low) level (the lack of smoothness of many of the curves arises from the integrality restriction on m and the selection of l, δ, and d from predefined discrete sets). The trends for Category B and C items are consistent with those for Category A items, and they are therefore omitted. In order to validate whether parameter values have a signicant impact on the bullwhip effect, we performed a series of hypothesis tests (using the Minitab statistical software package) on the difference in bullwhip effect values

Table 2-4. Computational results for category C with standard parameter settings.

Category C								
$\frac{d}{c} \times 100\%$	$\frac{\delta}{c} \times 100\%$	Q_1	q_1	Δ_R	Δ_S	Δ	BWE	$SW\%$
1%	4%	1527.17	320.31	0.08%	−0.59%	−0.14%	14.90	0.30%
2%	5%	2351.89	284.71	0.19%	−0.48%	−0.03%	14.51	0.62%
3%	5%	3148.86	259.39	0.33%	−0.61%	0.02%	16.38	0.87%
4%	6%	4023.69	211.73	0.53%	−0.66%	0.14%	15.28	1.33%
5%	6%	4835.66	185.04	0.76%	−1.06%	0.16%	16.48	1.62%
6%	6%	5241.2	172.36	1.04%	−1.60%	0.17%	17.04	1.76%
7%	6%	6051.4	148.34	1.37%	−2.27%	0.17%	18.11	2.04%
8%	6%	6860.45	126.06	1.74%	−3.09%	0.16%	19.13	2.32%
9%	7%	7822.08	63.87	2.17%	−3.69%	0.24%	17.59	3.03%
10%	7%	8644.95	41.79	2.64%	−4.80%	0.20%	18.44	3.35%
11%	7%	9821.55	376.66	3.17%	−6.33%	0.05%	20.02	3.66%
12%	7%	10641.8	358.34	3.76%	−7.77%	−0.03%	20.83	3.97%
13%	8%	11553.4	1018.19	4.40%	−9.18%	−0.06%	20.91	4.54%
14%	8%	12742.95	1354.38	5.10%	−11.25%	−0.26%	22.66	4.89%
15%	8%	13514.7	1700.77	5.87%	−13.27%	−0.41%	23.83	5.06%
16%	9%	14804.6	2712.24	6.70%	−15.54%	−0.60%	23.97	5.70%
17%	9%	15652.18	2694.44	7.60%	−17.77%	−0.73%	24.84	6.08%
18%	9%	16785.47	3396.47	8.57%	−20.72%	−1.05%	26.40	6.28%
19%	10%	18173.46	4041.66	9.61%	−23.59%	−1.29%	25.87	7.19%
20%	10%	18958.36	4388.96	10.73%	−26.54%	−1.50%	26.74	7.40%
21%	10%	20171.77	4730.3	11.93%	−30.15%	−1.89%	27.96	7.81%
22%	11%	21514.35	5735.9	13.21%	−33.89%	−2.25%	27.27	8.59%
23%	11%	22660.31	6439.43	14.57%	−38.11%	−2.72%	28.49	8.82%
24%	12%	24097.86	7082.9	16.02%	−42.35%	−3.14%	27.82	9.87%
25%	12%	25605.17	8142.26	17.56%	−47.81%	−3.90%	29.42	10.12%
26%	12%	26842.95	8487.74	19.20%	−52.81%	−4.44%	30.56	10.60%
27%	13%	28221.82	9490.93	20.94%	−58.01%	−4.98%	30.28	11.49%
28%	13%	29826.79	10193.54	22.78%	−64.40%	−5.84%	31.71	12.01%
29%	13%	30983.48	10900.95	24.72%	−70.29%	−6.47%	32.80	12.26%
30%	14%	32836.64	11901.21	26.78%	−77.30%	−7.39%	32.54	13.48%

Table 2-5. Results of computational tests.

	Category	Δ_R	Δ_S	Δ	BWE
Average	A	16.22%	−13.33%	0.12%	13.32
	B	12.21%	−16.92%	−0.79%	18.33
	C	8.80%	−22.86%	−1.59%	23.43
Maximum	A	50.86%(30%)	0.37%(2%)	1.28%(13%)	17.20(30%)
	B	37.72%(30%)	0.25%(2%)	0.70%(12%)	25.18(30%)
	C	26.78%(30%)	−0.48%(2%)	0.24%(9%)	32.54(30%)
Minimum	A	0.10%(1%)	−50.01%(30%)	−4.11%(30%)	9.22(2%)
	B	0.09%(1%)	−60.71%(30%)	−6.22%(30%)	11.57%(2%)
	C	0.08%(1%)	−77.30%(30%)	−7.39%(30%)	14.90(1%)

Table 2-6. Parameter settings for factorial tests.

Category	Level	A	i	r_f	γ
A	high	800	0.5	500	ln(1.66)
	low	40	0.125	0	0
B	high	100	0.5	4000	ln(1.66)
	low	5	0.125	0	0
C	high	20	0.5	2000	ln(1.66)
	low	1	0.125	0	0

Table 2-7. Hypothesis tests for bullwhip effect with key parameters at different levels.

Parameter	Alternative hypothesis	P-value
$H_0 : BWE^H \geq BWE^L$		
A	$H_1 : BWE^H < BWE^L$	0
i	$H_1 : BWE^H < BWE^L$	0
γ	$H_1 : BWE^H < BWE^L$	0
r_f	$H_1 : BWE^H < BWE^L$	0

for high and low parameter settings. That is, we computed the pairwise differences in bullwhip effect for each problem instance with low (L) and high (H) parameter value settings, and performed a hypothesis test to determine whether the difference was statistically different from zero. The results of these hypothesis tests, which demonstrate that the bullwhip effect is lower at high parameter value settings, are shown in Table 2-7.

The results shown in the figures and hypothesis tests lead to the following observations.

For all inventory categories, a higher value of the retailer's set-up cost, A, implies smaller values of Δ_R and the BWE. This is because a higher value of A induces a longer regular replenishment period length, t_0. Because we require $Q = mt_0$ with m integer, it follows that a larger value of t_0 restricts the retailer's flexibility in setting Q, which contributes to the decrease of Δ_R in A.

For all inventory categories, a higher value of the retailer's per unit holding cost rate, i, implies smaller values of Δ_R and the BWE. Although a higher value of i implies a smaller t_0, the main effect of a higher value of i is that it discourages the retailer from

placing large orders. This effect reduces the bullwhip effect and hinders the retailer's potential to gain additional revenue from a longer promotion period length.

For all inventory categories, a higher value of $\tilde{r}_i$ (the number of units purchased per dollar of price reduction for impulsive consumers) implies a higher value Δ and a lower value of the BWE. This is because the presence of impulsive consumers makes it more likely for the retailer to pass a larger percentage of the supplier's discount to consumers instead of forward buying. Consequently, the retailer's order pattern is more consistent with the end consumers' purchase pattern, which leads to a smaller degree of bullwhip effect.

For all inventory categories, a higher value of $\tilde{r}_f$ (the number of units purchased per dollar of price reduction for brand loyals) leads to a lower value of Δ, which indicates that if a retailer price discount induces a large degree of consumer forward buying, the supplier should adopt a conservative promotion strategy.

For all inventory categories, the bullwhip effect exists and, except for the results when γ is at a low level, the BWE tends to increase as d increases; when γ is at a low level, the BWE decreases. To better understand this phenomenon, Figure 2-4 illustrates how the ratio of the retailer's forward buy quantity to the retailer's special order quantity, $\frac{q_1}{Q_1}$, changes with d. We observe that when γ is at a low level, the BWE is consistent with the value of $\frac{q_1}{Q_1}$. This indicates that the supplier discount, d, induces the bullwhip effect via the interaction of two quantities: Q_1 and q_1. As we mentioned before, when presented with a price discount, the retailer will place a special order of size $Q_1 > q_0$, and a larger value of d clearly leads to a larger value of Q_1. A larger value of Q_1, in turn leads to a larger value of $std(Q)$. As part of its large order, the retailer may also adopt a forward buying strategy, i.e., $q_1 > 0$. A forward buy quantity increases the variance of orders placed with the supplier, but does not stimulate consumer demand. Thus, the larger the ratio of $\frac{q_1}{Q_1}$, the larger the bullwhip effect; the bullwhip effect is then maximal when $\frac{q_1}{Q_1} = 1$ (which implies $std(D) = 0$ and $\frac{std(Q)}{std(D)} = \infty$), and minimal when $q_1 = 0$.

Consequently, when γ is at a low level, q_1 is significantly large, so that the impact of q_1 overwhelms the impact of Q_1; as a result, the bullwhip effect changes in response to the change in $\frac{q_1}{Q_1}$. No systematic relationship can be observed between the bullwhip effect

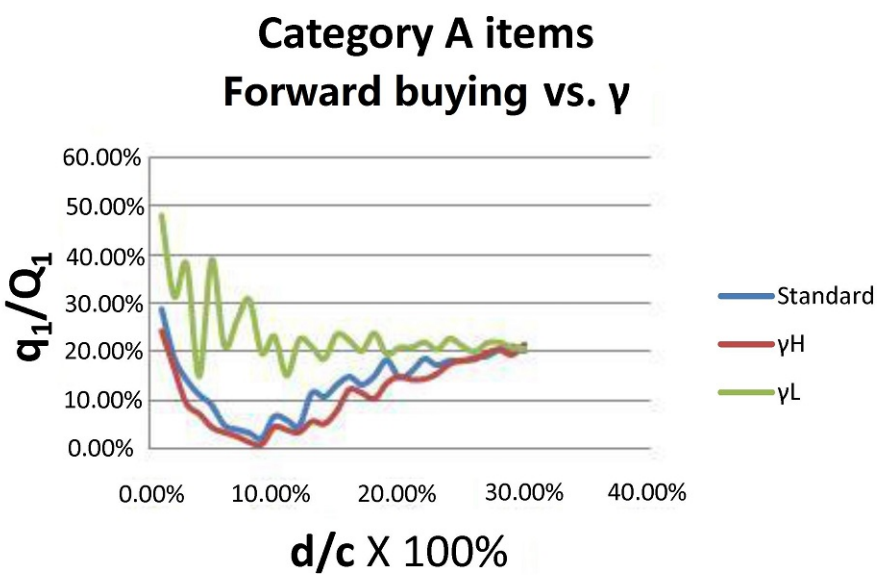

Figure 2-4. Forward buying in supplier discount with γ at different levels.

and Δ_R, Δ_S and Δ. In other words, a higher level of bullwhip effect does not universally imply lower levels of profit.

In summary, we demonstrated that even though the use of trade promotions can indeed increase a supplier's operations costs, these costs may be more than offset by increased revenues, even in the absence of explicit coordination. That is, the supply chain profit can exceed that under an EDLP strategy, if (i) the supplier judiciously sets the price discount; (ii) there is a sufficient number of impulsive consumers who buy the product at the discounted price; (iii) the price discount does not induce a high degree of end-consumer forward buying.

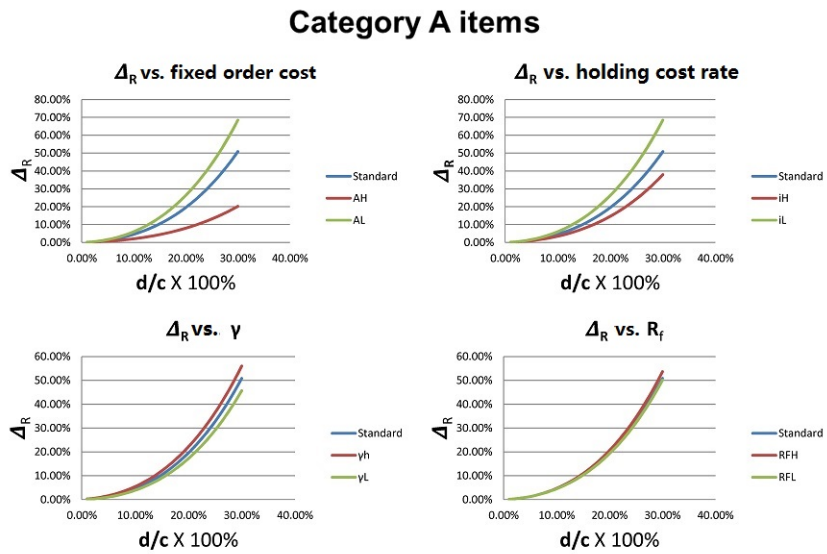

Figure 2-5. Retailer profit gain in supplier discount with key parameters at different levels.

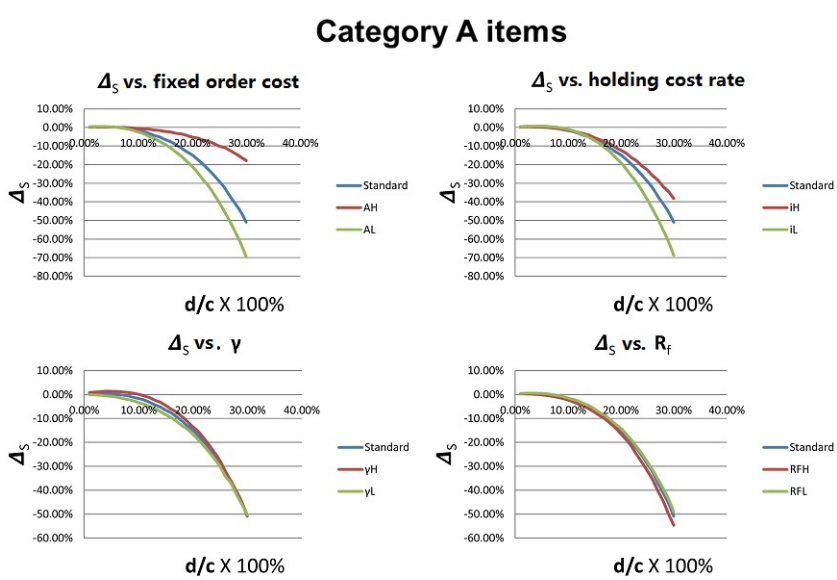

Figure 2-6. Supplier profit gain in supplier discount with key parameters at different levels.

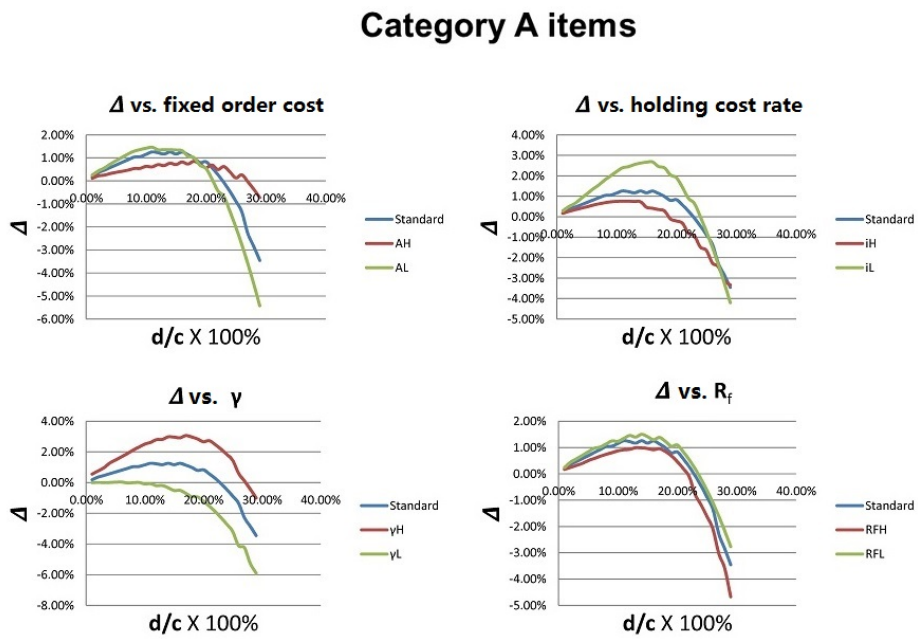

Figure 2-7. System profit gain in supplier discount with key parameters at different levels.

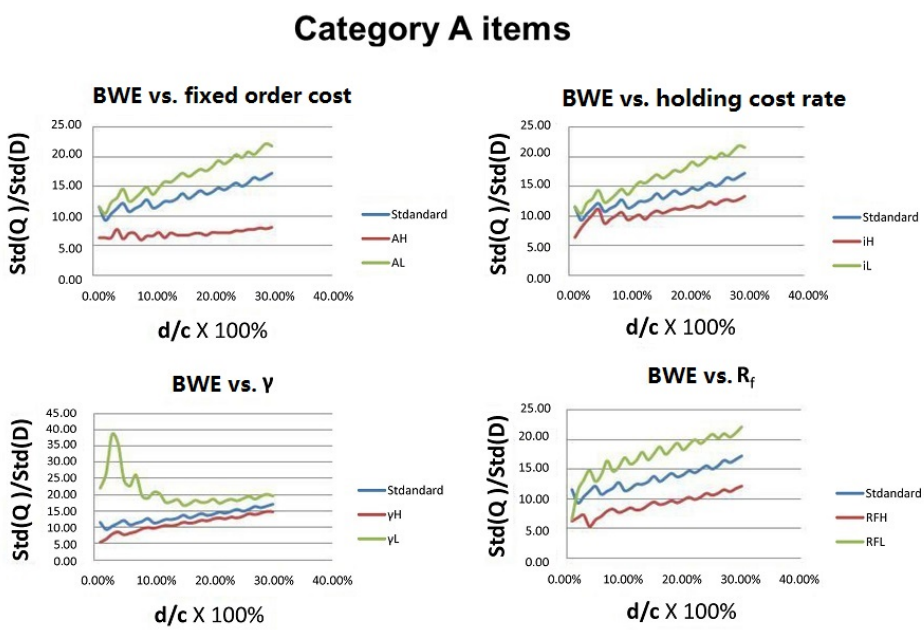

Figure 2-8. Bullwhip effect in supplier discount with key parameters at different levels.

CHAPTER 3
MULTI-PERIOD PRICE PROMOTIONS IN A SINGLE-SUPPLIER, MULTI-RETAILER SUPPLY CHAIN UNDER ASYMMETRIC DEMAND INFORMATION

3.1 Problem Description and Literature Survey

A study by Andersen Consulting [42] found that "trade promotion is the biggest, most complex and controversial dilemma facing the retail industry today." As we mentioned in Chapter 1, trade promotions have long played an important role in U.S. retail supply chains, however, trade promotions do not effectively improve suppliers' performance. Adoption of "off-invoice" deals is one of the major reasons for the poor performance of trade promotions, because it is profitable for retailers to respond to an off-invoice trade promotion by engaging in *forward-buying* and *diverting*, and these forward-buying and diverting strategies result in the bullwhip effect, which is viewed in extremely negative terms because of its negative impacts on supply chain operations costs. Lee, Padmanabhan and Wang [64] characterized price variations as one of the major causes of the bullwhip effect and suggested using corresponding management practices (an EDLP strategy, for example) to reduce the operations costs associated with the bullwhip effect.

Despite the many undesirable consequences discussed in the operations literature, there has been no sign of decline in the use of trade promotions in industry. Some marketing science literature shows that a carefully designed price promotion scheme can improve the supplier's profit when compared to the case of everyday low pricing (EDLP). Blattberg, Eppen and Lieberman [25] suggested that effective timing of off-invoice trade promotions can reduce a supplier's inventory cost. Drèze and Bell [42] showed that scan-back trade promotions may improve both the retailer's and the supplier's performance. Scan-back trade promotions are so named because they are based on stores' scanner data and, therefore, the amount sold to end consumers from each store.

Although the marketing science literature provides important insights about trade promotions, they typically do not explicitly consider supply chain operations costs. Mathematical models that account for operations costs in addition to the demand-side effects of price promotions have been developed in the operations research literature. Neslin, Powell and Stone [77] developed a multi-echelon model that considers the actions of a supplier, a set of retailers, and consumers from the point of view of the supplier, who attempts to maximize profit by targeting advertising to consumers and trade promotions to retailers. Due to the complexity of their model, they could not provide a procedure for obtaining a globally optimal solution. Moreover, their modeling approach parameterized on all of the retailer's decisions. However, trade promotions are supplier decisions, which affect the retailer's actions, and there are numerous interactions taking place between the supplier and the retailer. Usually in these interactions, the supplier is the leader who sets the wholesale prices and discounts first, and the retailer is the follower who sets the retail prices and places purchase orders given the supplier's decisions. As a result, a Stackelberg game in which the supplier moves first is a better approximation of this process. Kopalle, Mela and Marsh [62] were among the first who developed a normative model using a Stackelberg game to determine optimal supplier and retailer prices over time. In their study, they indicated that trade promotions have effects not only in the periods they are offered, but also affect future periods. As a result, they developed a dynamic, descriptive brand sales model which accounts for dynamic effects of discounts, such as consumer forward buying and competition between brands, and they integrated this descriptive model with the normative model to study the contemporaneous effects and future effects of promotions. However, their model allowed no retailer forward buying and diverting. Kogan and Herbon [60] considered a similar two-echelon supply chain model with a supplier and a retailer facing stochastic demands, and they used a dynamic continuous-time Stackelberg game to simulate the interactions between the supplier and he retailer.

In contrast to the study of Kopalle et al. [62], in their model, Kogan and Herbon [60] incorporated retailer forward buying and considered instantaneous and exogenous changes change in the consumer price sensitivity which will temporarily increase the consumer demand potential. In addition to these two studies, there are many other studies about channel coordination dealing with pricing strategies for a supplier-retailer channel relationship in the literature (e.g., Jeuland and Shugan [59]; Eliashberg and Steinberg [44]; Xie and Wei [100]).

In this paper, we provide a mathematical model of a decentralized two-echelon supply chain in which the supplier's pricing decisions (trade promotion levels) and the retailer's operations decisions (order quantities, inventory levels, and transshipment quantities) are determined simultaneously. To broaden the applicability of our model, we assume that the price can be dynamically changed over time and that demand is stochastic and price-dependent. To the best of our knowledge, no existing work considers an exact solution method for a joint promotion and operations problem in a multi-echelon supply chain under uncertain demand, as we do in this chapter. In the field of mathematical programming, this problem falls in the class of linearly constrained, bilevel, nonconvex optimization problems. Most of the available algorithms in the field of bilevel programming apply to bilevel linear problems (where, for fixed values of one set of decision variables, the remaining problem becomes a linear program). Ben-Ayed [21] and Wen and Hsu [98] provided detailed reviews of bilevel linear programming problems. They presented a basic model along with characterizations of optimal solution properties for the problem class and some existing solution approaches. To solve linearly constrained bilevel convex quadratic problems, Muu and Van Quy [74] developed a branch-and-bound algorithm for finding a global optimal solution. As we will see, none of the previously mentioned solution methods for bilinear optimization problems can be directly applied to the model we develop, which falls within a more general class of nonconvex bilevel optimization problems. These methods have, however, provided

substantial guidance and inspiration for the solution procedure we have developed for solving the problem we define.

As we later discuss, our model can be cast as a generalized bilinear program (GBP). Al-Khayyal [7] provided a generic approach for solving GBPs globally; this approach was essentially an extension to the Al-Khayyal and Falk [9] algorithm which was proposed to solve jointly constrained bilinear programs. Al-Khayyal, Larsen and Van Voorhis [8] further extended the applicability of this method to nonconvex quadratically constrained quadratic programs. Our solution approach adapts the branch-and-bound algorithm of Al-Khayyal [7] and also draws on previously developed Reformulation-Linearization (RLT) techniques. Sherali and Alameddine [84] developed a branch-and-bound algorithm based on an RLT for jointly constrained bilinear programs. Although the linear relaxation obtained from the RLT approach is theoretically tighter than that derived by Al-Khayyal's method, this RLT branch-and-bound approach cannot be applied to GBPs directly. In our study, we develop a method which is able to solve our problem by integrating the RLT branch-and-bound algorithm within a penalty-function-based approach, which penalizes violations of a relaxed constraint set in the objective function.

The remainder of this chapter is organized as follows. Section 3.2 describes our supply chain model. We present three approaches for solving the resulting problem in Section 3.3. We first transform our original model to a generalized bilinear programming problem in Sections 3.3.1 and 3.3.2. Section 3.3.3.1 presents a widely used heuristic solution method previously developed for solving bilinear problems, followed by Sections 3.3.3.2 and 3.3.3.3, which present two exact solution methods. The first of these is a modification of Al-Khayyal's approach [7] for solving generalized bilinear programs and the second is a new approach we have developed for this problem class. In Section 3.4, we discuss the results of a computational study used to validate our solution method and compare it with the results of the heuristic method and Al-Khayyal's approach.

3.2 Problem Statement and Formulation

To formalize our model, we define the following notation:

Inputs and Parameters

i, j: retail store indices, $i, j = 1, \ldots, S$.

l: period index, $l = 1, \ldots, L$.

L^P : length of promotion period, $1 < L^P < L$.

c_i^l: supplier's initial unit wholesale price for store i in period l.

m_i^l: supplier's unit production and transportation cost for retail store i in period l.

h_i^l: retailer's unit inventory holding cost at store i in period l.

t_{ij}^l: retailer's unit transportation cost between stores i and j in period l, where $i \neq j$.

d_i^l: base value of deterministic component of consumer demand at store i in period l.

α_i^l: retailer's pass-through rate at store i in period l.

γ_i^l: consumer promotional price elasticity at store i in period l.

Decision Variables

z_i^l: supplier's price discount at store i in period l.

x_i^l: retailer's order quantity at store i in period l.

I_i^l: retailer's inventory level at store i at the end of period l.

y_{ij}^l: retailer's transhipment quantity from store i to store j in period l.

We consider the following two-echelon system. A supplier distributes a single product to a set of retail stores who in turn sell the item to consumers over a finite planning horizon of L periods. The retailer runs a chain of S stores, which serve geographically dispersed, heterogeneous markets, and face random, price-dependent demand at each store.

The supplier and retailer act non-cooperatively and in a sequential manner. The supplier (upper level) selects trade promotion level z_i^l targeted at each retail store in each period in order to maximize its expected profit. Given the supplier's price discounts, the retailer (lower level) optimizes its profit by choosing: (1) the order quantity at store i

in period l, denoted by x_i^l; (2) the inventory level at store i at the end of period l, denoted by I_i^l; and (3) the diverting quantity from store i to store j in period l, denoted by y_{ij}^l, for all periods $l = 1, \dots, L$ and all stores $i, j = 1, \dots, S$, $i \neq j$. The supplier explicitly incorporates the anticipated reactions of the retailer in its optimization process, and we assume full information is available to the supplier regarding the retailer's inventory and transportation costs. This fits the classical Stackelberg game paradigm, where the leader, who is aware of the follower's best response, chooses a move that maximizes its own expected payoff.

We assume randomness in demand is price-independent and can be modeled in an additive fashion. Specifically, demand is defined as $\tilde{d}_i^l(\omega) = d_i^l + \alpha_i^l \gamma_i^l z_i^l + \varepsilon_i^l(\omega)$, where ω is an outcome belonging to some probability space Ω, and ω influences all random variables ε_i^l. Note that the quantity $\alpha_i^l z_i^l$ is the price discount seen by consumers as it is the product of the retailer's discount pass-through rate and the supplier's trade discount. Thus, the quantity $\alpha_i^l \gamma_i^l z_i^l$ corresponds to the increase in demand at retail store i in period l as a result of the discount passed through to consumers.

As noted in Section 3.1, we assume that the retailer makes its decisions (x_i^l, I_i^L, and y_{ij}^l) after demand uncertainty is resolved but before demand occurs. One interpretation of this assumption is that, from a marketing perspective, retailers are, by definition, closer to consumers than manufacturing companies, and so retailers can more easily engage in personal contact with consumers, gather information on consumer behavior, and anticipate consumer purchase patterns in the short run. And, because retailer-to-supplier lead times are often very short, the retailer has the luxury of placing orders immediately before demand occurs (or even in response to demand in some cases, as in a make-to-order setting). Moreover, effective promotions are usually those offered only over a short time-span, because frequent usage of promotions diminishes their effectiveness [26]. As a result, we can often assume that a retailer can accurately forecast consumer demand during promotion periods in the near future.

However, we assume that the retailer does not share consumer information with the supplier in this decentralized supply chain. Because the supplier does not possess the retailer's specialized knowledge of local markets, the supplier cannot accurately forecast consumers demand at the local level; as a result, the supplier must determine its promotional discount levels (z_i^l values) in the absence of precise local demand information. Thus, the asymmetry in the degree of demand uncertainty between the supplier and retailer is attributable to the retailer's ability to obtain more precise local demand information. From the supplier's perspective it is either too costly to obtain this local information or the retailer is unwilling to share this information for competitive reasons.

Our model captures the interaction between the supplier and retail chain. The purpose of the model is to determine the optimal promotion plan for the supplier when anticipating the retail chain's usage of forward buying and diverting strategies. To simplify the exposition of the model and for model tractability, we assume that: 1) replenishment delays are negligible; 2) fixed costs are zero (or constant); 3) transhipment between any two stores can be done within one period; and 4) the retailer's discount pass-through rates are fixed or pre-determined[1] . Based on the above notation and model description, the trade promotion problem can be formulated as a stochastic bilevel program with bilinear objectives at both decision levels, and with linear constraints. For a given realization ω, the order quantity $x_i^l(\omega)$, forward buying quantity $I_i^l(\omega)$, and diverting quantity $y_{ij}^l(\omega)$ correspond to an optimal solution to the lower level linear program for given upper level trade promotion levels z_i^l, the optimal values of

[1] While we recognize that the pass-through rates are decision variables set by the retailer, we assume fixed values in our model for model tractability. Thus, when using this model for decision making, the supplier must parameterize on the pass-through rate in order to determine (at least approximately) the values that the retailer is likely to apply.

which are determined by maximizing the expected value of net supplier profit across all possible realizations.

We can now formulate our stochastic multi-period, two-stage trade promotion problem as:

$$\text{(STP)} \quad \max_{z} \mathbb{E}_{\xi}\left[\sum_{l=1}^{L}\sum_{i=1}^{S}\left(c_i^l - m_i^l - z_i^l\right) x_i^l(\omega)\right] \tag{3–1}$$

$$\text{s.t. } 0 \le z_i^l \le c_i^l - m_i^l \qquad \forall i = 1, \dots, S, l = 1, \dots, L^P \tag{3–2}$$

$$z_i^l = 0 \qquad \forall i = 1, \dots, S, l = L^P + 1, \dots, L \tag{3–3}$$

where $(x(\omega), I(\omega), y(\omega))$ is an optimal solution of the following problem:

$$\min_{x,y,I} \sum_{l=1}^{L}\sum_{i=1}^{S}\left[\left(c_i^l - z_i^l\right) x_i^l(\omega) + h_i^l I_i^l(\omega) + \sum_{j\neq i} t_{ij}^l y_{ij}^l(\omega)\right] \tag{3–4}$$

$$\text{s.t. } I_i^l(\omega) = I_i^{l-1}(\omega) + x_i^l(\omega) + \sum_{k\neq i} y_{ki}^{l-1}(\omega) - \tilde{d}_i^l(\omega) - \sum_{j\neq i} y_{ij}^l(\omega)$$

$$\forall i = 1, \dots, S, l = 1, \dots, L \tag{3–5}$$

$$\tilde{d}_i^l(\omega) = d_i^l + \alpha_i^l \gamma_i^l z_i^l + \varepsilon_i^l(\omega) \qquad \forall i = 1, \dots, S, l = 1, \dots, L \tag{3–6}$$

$$x_i^l(\omega) \ge 0, I_i^l(\omega) \ge 0 \qquad \forall i = 1, \dots, S, l = 1, \dots, L$$

$$y_{ij}^l(\omega) \ge 0 \qquad \forall i = 1, \dots, S, i \neq j, l = 1, \dots, L$$

$$I_i^0(\omega) = 0 \qquad \forall i = 1, \dots, S \ ,$$

where $\xi(\omega)$ is a random vector consisting of all random components $\varepsilon_i^l(\omega)$. The upper level objective (3–1) is to maximize expected net profit and is expressed as the difference between the sum of revenues arising from wholesale pricing $(c_i^l - z_i^l)$ and the sum of variable costs. The upper level constraints (3–2) state that the wholesale price discounts z_i^l should be nonnegative and less than or equal to the per unit profit margin, $c_i^l - m_i^l$. The upper level constraints (3–3) state that, after the promotion period, the supplier stops offering a trade promotion to all retail stores, and we add this set of constraints to our model so that it is able to capture the carry-over effect (forward

buying, for example) of decisions in promotion periods. The objective of the lower level problem (3–4) is to minimize the retailer's total cost of ordering, holding inventory and diverting. In our study, since discount pass-through rates are fixed, it follows that the retail prices are indirectly determined by the supplier. As a result, for the lower level problem, minimizing cost is equivalent to maximizing profit. The first set of lower level constraints (3–5) requires that ending inventory at store i in period l equals the ending inventory at store i in period $l-1$, plus the amount ordered from the supplier at store i in period l, plus the amount shipped from other stores to store i in period $l-1$, minus the demand at store i in period l, minus the amount shipped from store i to other stores in period l. The second set of constraints (3–6) models consumer demand as a linear function of the wholesale price discount. The remaining constraints indicate that all variables are real-valued and nonnegative.

Note that for a realization ω and specific promotion levels, the objective of the lower level problem is convex (but not strictly convex), which implies the solution to the lower level problem may not necessarily be unique. As a result, we assume that given the choice between solutions to the lower level problem with equal cost, the solution selected is the one yielding the highest expected profit for the supplier. One may alternatively apply a worst-case approach from the supplier's perspective, assuming that the retailer chooses the solution yielding the lowest expected profit for the supplier.

There are two difficulties inherent in solving STP. First, for each outcome of the demand realization, the resulting problem is a bilevel problem with bilinear objectives at both levels and with linear constraints, which falls in the class of NP-hard bilinearly constrained bilinear programs (or generalized bilinear programs). The second difficulty lies in the computational burden of computing the expected profit function.

In the next section, we propose a three step methodology to solve the STP problem globally. First, the stochastic trade promotion problem is converted to an equivalent deterministic problem. Then, we transform the resulting deterministic bilevel problem

into a single-level problem. Finally, we adapt a branch-and-bound algorithm based on a Reformulation-Linearization Technique (RLT) for solving the resulting single level generalized bilinear program.

3.3 Solution Procedures for STP

3.3.1 Deterministic Equivalent

In order to analyze the STP, we first construct a supply chain *time-expanded network* $G = \{N, A\}$ representing the two-echelon, multi-period problem, where $N = \{(i, l) : i = 1, \dots, S,\ \ l = 1, \dots, L\} \cup \{(0, 0)\}$ is the set of nodes and A is the set of arcs. Node $(0, 0)$ represents the supplier, while node (i, l) corresponds to retail store i in period l. The set $N^l = \{(1, l), (2, l), \dots, (S, l)\} \subset N$ corresponds to the set of nodes associated with retail stores in period l. Three types of arcs are contained in the network G: (1) $((0, 0), (i, l)), \forall (i, l) \in N - \{(0, 0)\}$, with arc cost $c_i^l - z_i^l$ per unit of flow (order quantity) from the supplier to node (i, l); (2) $((i, l), (i, l + 1)), \forall (i, l) \in N - N^L - \{(0, 0)\}$ with a unit cost of h_i^l corresponding to the inventory carried from node (i, l) to $(i, l + 1)$; and (3) $((i, l), (j, l + 1)), \forall i, j = 1, \dots, S, l = 1, \dots, L - 1, i \neq j$ with a unit cost of t_{ij}^l for transshipment from node (i, l) to node $(j, l + 1)$.

From the construction of graph G, we might view the lower level problem of STP as a minimum cost flow problem, which requires sending d_i^l units of flow as cheaply as possible from node $(0, 0)$ to each node (i, l) in the set $N - \{(0, 0)\}$ in an uncapacitated network. Observe that a minimum cost flow problem with no arc capacities can be decomposed into a set of $S \times L$ shortest path problems that are independent of the demand levels (Appendix C for a demonstration of the validity of this decomposition). This observation implies that the lower level problem of STP is equivalent to a set of deterministic problems that are independent of the random terms. Based on this observation, it is not hard to see that for the entire problem, the stochastic components only appear in the objective function of the top-level problem. After taking the expectation of the objective function, the STP is thus a deterministic problem, and

the distributions of the stochastic components do not impact the optimization problem formulation. We next discuss how to transform this deterministic equivalent for the two-level problem into a single-level optimization problem.

3.3.2 Single-level Problem

The deterministic trade promotion problem (DTP) is derived directly from the STP problem by replacing the random term with its expectation. The results of Section 3.3.1 imply that DTP is equivalent to STP. Note that the lower level problem can be written without holding costs by making substitution $I_i^l = \sum_{\tau=1}^{l}[x_i^\tau + \sum_{k\neq i} y_{ki}^{\tau-1} - \sum_{j\neq i} y_{ij}^\tau - (d_i^\tau + \alpha_i^\tau \gamma_i^\tau z_i^\tau)]$; then the DTP problem can be formulated as:

$$
\begin{aligned}
(\text{DTP}) \ \max_{x,y,z} \ & \sum_{l=1}^{L}\sum_{i=1}^{S} \left(c_i^l - m_i^l - z_i^l\right) x_i^l \\
\text{s.t. } & 0 \leq z_i^l \leq c_i^l - m_i^l && \forall i = 1, \dots, S, l = 1, \dots, L^P \\
& z_i^l = 0 && \forall i = 1, \dots, S, l = L^P + 1, \dots, L \ ,
\end{aligned}
$$

where (x, y) is an optimal solution of

$$
\begin{aligned}
\min_{x,y} \ & \sum_{l=1}^{L}\sum_{i=1}^{S} \left[\left(\bar{c}_i^l - z_i^l\right) x_i^l + \sum_{j\neq i} \bar{t}_{ij}^l y_{ij}^l - \bar{h}_i^l \left(d_i^l + \alpha_i^l \gamma_i^l z_i^l\right) \right] \\
\text{s.t. } & \sum_{\tau=1}^{l}(x_i^\tau + \sum_{k\neq i} y_{ki}^{\tau-1} - \sum_{j\neq i} y_{ij}^\tau) \geq \sum_{\tau=1}^{l}(d_i^\tau + \alpha_i^\tau \gamma_i^\tau z_i^\tau) \quad \forall i = 1, \dots, S, l = 1, \dots, L \\
& x_i^l \geq 0 && \forall i = 1, \dots, S, l = 1, \dots, L \\
& y_{ij}^l \geq 0 && \forall i = 1, \dots, S, i \neq j, l = 1, \dots, L \ ,
\end{aligned}
\tag{3–7}
$$

where $\bar{h}_i^l = \sum_{\tau=l}^{L} h_i^\tau$, $\bar{c}_i^l = c_i^l + \bar{h}_i^l$ and $\bar{t}_{ij}^l = t_{ij}^l + \bar{h}_j^{l+1} - \bar{h}_i^l$. For any fixed value of z_i^l, the lower level problem is a linear program, so any optimal solution of the lower level problem satisfies the strong duality property. As a result, we can replace the lower level program by its primal-dual optimality conditions, where π is the vector of dual variables

associated with the set of constraints (3–7). Let us first define:

$$F \equiv \left\{ (x,y,z,\pi): \begin{array}{ll} \sum_{\tau=1}^{l}(x_i^\tau + \sum_{k\neq i} y_{ki}^{\tau-1} - \sum_{j\neq i} y_{ij}^\tau - \alpha_i^\tau\gamma_i^\tau z_i^\tau) \geq \sum_{\tau=1}^{l} d_i^\tau, & \forall i = 1,\dots,S, \\ & l = 1,\dots,L \\ \sum_{\tau=l}^{L} \pi_i^\tau + z_i^l \leq \bar{c}_i^l, & \forall i = 1,\dots,S, \\ & l = 1,\dots,L \\ \sum_{\tau=l+1}^{L} \pi_j^\tau - \sum_{\tau=l}^{L} \pi_i^\tau \leq \bar{t}_{ij}^l, & \forall i = 1,\dots,S, i \neq j, \\ & l = 1,\dots,L \end{array} \right\}$$

and

$$\Omega \equiv \left\{ (x,y,z,\pi): \begin{array}{ll} z_i^l \geq 0 \text{ and } c_i^l - m_i^l - z_i^l \geq 0, & \forall i = 1,\dots,S, l = 1,\dots,L^P \\ z_i^l = 0, & \forall i = 1,\dots,S, l = L^P,\dots,L \\ x_i^l \geq 0 \text{ and } U^l - x_i^l \geq 0, & \forall i = 1,\dots,S, l = 1,\dots,L \\ y_{ij}^l \geq 0 \text{ and } U^l - y_{ij}^l, \geq 0, & \forall i = 1,\dots,S, i \neq j, l = 1,\dots,L \\ \pi_i^l \geq 0 \text{ and } \bar{c}_i^l - \pi_i^l \geq 0, & \forall i = 1,\dots,S, l = 1,\dots,L \end{array} \right\} ,$$

where $U^l = \sum_{\tau=l}^{L}\sum_{i=1}^{S}(d_i^\tau + \alpha_i^\tau\gamma_i^\tau m_i^\tau)$ and Ω is a hyper-rectangle. Using the above definitions we can formulate the DTP as follows:

$$(\text{SDTP}(\Omega)) \quad \max_{x,y,z,\pi} \phi(x,y,z,\pi) = \sum_{l=1}^{L}\sum_{i=1}^{S}\left(c_i^l - m_i^l - z_i^l\right) x_i^l \tag{3–8}$$

$$\text{s.t.} \quad \sum_{l=1}^{L}\sum_{i=1}^{S}\left[\left(\bar{c}_i^l - z_i^l\right) x_i^l + \sum_{j\neq i} \bar{t}_{ij}^l y_{ij}^l\right] = \sum_{l=1}^{L}\sum_{i=1}^{S}\sum_{\tau=l}^{L} \pi_i^\tau \left(d_i^l + \alpha_i^l\gamma_i^l z_i^l\right) \tag{3–9}$$

$$(x,y,z,\pi) \in F \cap \Omega \ .$$

The above problem maximizes a bilinear objective function (3–8) over a feasible region defined by a bilinear constraint (3–9) and a set of linear constraints. This problem thus falls within the class of generalized bilinear programs (GBPs), and so the SDTP reduces to a linear program whenever either z or (x,y,π) are fixed. However, the objective function (3–8) and the bilinear constraint (3–9) are nonconvex functions. Solving a GBP

is NP-hard [79], and, to the best of our knowledge, no methods exist that guarantee convergence to an exact solution in a finite number of steps. In the following section, we will present some classical techniques for solving the SDTP problem, as well as a penalty-based method that we have developed for solving the problem.

3.3.3 Linearization

3.3.3.1 Successive linear programming approach

Successive linear programming is a commonly used heuristic method for solving bilinear programming problems. This procedure iterates between fixing the supplier's price discounts z and the retailer's primal and dual variables (x, y, π) for solving SDTP. At a given iteration k, we first find the values of (x^k, y^k, π^k) that optimize the objective function for a fixed z^{k-1}, and then find the vector z^k that optimizes the objective function for fixed values of (x^k, y^k, π^k). We repeat this procedure until the objective does not improve between two successive iterations.

The classical bilinear program is a class of quadratic programs with the following structure:

$$\begin{aligned} \max_{x,y} \ & c^T x + x^T A y + d^T y \\ \text{s.t. } & x \in X := \{x : B_1 x \leq b_1, x \geq 0\} \\ & y \in Y := \{y : B_2 y \leq b_2, y \geq 0\} \ . \end{aligned}$$

If the above bilinear program has a finite optimal solution, then there exists an extreme point $x^* \in X$ and an extreme point $y^* \in Y$ such that (x^*, y^*) is an optimal solution of the classical bilinear program [96]. Since the feasible region is defined by two separable polyhedral sets, the classical bilinear program is also called a bilinear program with separable constraints. Sherali and Shetty [85] showed that, for a classical bilinear program, the limit point of the successive linear programming approach is a locally optimal solution. However, SDTP is a GBP problem, which does not have separable

constraints in the bilinear terms. Therefore, we have no guarantees on solution quality for a solution obtained using successive linear programming approach.

3.3.3.2 Al-Khayyal's approach

Al-Khayyal and Falk [9] developed an infinitely convergent branch and bound algorithm for jointly constrained bilinear programs (JCBPs) using lower bounds derived from convex envelopes of the bilinear terms. Al-Khayyal [7] found that the same approach can also be applied for GBPs.

The first step of the approach is to obtain the concave overestimate of the objective function (3–8) over Ω. Piecing together all the variables, we obtain a vector $\Lambda \equiv (x, y, z, \pi, \eta, \zeta, \nu)$, with up to $N = (5 + S + L) \times (S \times L)$ components, and the concave overestimate of (3–8) over Ω can be represented as:

$$\psi(\Lambda) = \sum_{l=1}^{L} \sum_{i=1}^{S} \left[(c_i^l - m_i^l) x_i^l - \eta_i^l \right]$$
$$\text{s.t. } \Lambda \in F_1(\Omega) \ , \tag{3–10}$$

where

$$F_1(\Omega) \equiv \left\{ \Lambda : \begin{array}{cc} \eta_i^l \geq 0 & \forall i = 1, \dots, S, l = 1, \dots, L \\ \eta_i^l \geq (c_i^l - m_i^l) x_i^l + U^l z_i^l - (c_i^l - m_i^l) U^l & \forall i = 1, \dots, S, l = 1, \dots, L \end{array} \right\} .$$

The second step of the approach is to obtain a polyhedral approximation of the convex hull of the region defined by constraint (3–9).

By the weak duality theorem, whenever (x, y) and π are feasible for the primal and dual problems, respectively, the left-hand-side of equation (3–9) is always greater than or equal to its right-hand-side. As a result, replacing constraint (3–9) with the following constraint does not change the feasible region of the SDTP:

$$\sum_{l=1}^{L} \sum_{i=1}^{S} (\bar{c}_i^l x_i^l + \sum_{j \neq i} \bar{t}_{ij}^l y_{ij}^l - \sum_{\tau=1}^{l} d_i^\tau \pi_i^l - z_i^l x_i^l - \sum_{\tau=1}^{l} \alpha_i^\tau \gamma_i^\tau z_i^\tau \pi_i^l) \leq 0 \ . \tag{3–11}$$

Next, define the polyhedral set

$$F_2(\Omega) \equiv \left\{ \Lambda : \begin{array}{ll} \sum_{i=1}^{S} \sum_{l=1}^{L} (\zeta_i^l + \sum_{\tau=1}^{l} \nu_i^{\tau l}) \leq 0 & \\ (\bar{c}_i^l - c_i^l + m_i^l) x_i^l + \sum_{j \neq i} \bar{t}_{ij}^l y_{ij}^l \leq \zeta_i^l & \forall i = 1, \dots, S, l = 1, \dots, L \\ \bar{c}_i^l x_i^l + \sum_{j \neq i} \bar{t}_{ij}^l y_{ij}^l - U^l z_i^l \leq \zeta_i^l & \forall i = 1, \dots, S, l = 1, \dots, L \\ -[d_i^\tau + \alpha_i^\tau \gamma_i^\tau (c_i^\tau - m_i^\tau)] \pi_i^l \leq \nu_i^{\tau l} & \forall i = 1, \dots, S, l = 1, \dots, L, \tau \leq l \\ -d_i^\tau \pi_i^l - \alpha_i^\tau \gamma_i^\tau \bar{c}_i^l z_i^\tau \leq \nu_i^{\tau l} & \forall i = 1, \dots, S, l = 1, \dots, L, \tau \leq l \end{array} \right\} .$$

Al-Khayyal [7] showed that for any (x, y, z, π) satisfying constraint (3–9), there exists $\Lambda \in F_2(\Omega)$.

From the results above, we have the following two observations: i) $\forall (x, y, z, \pi) \in F \cap \Omega$ and $\Lambda \in F \cap \Omega \cap F_1(\Omega)$, we have $\phi(x, y, z, \pi) \leq \psi(\Lambda)$; and ii) for any feasible solution (x, y, z, π) to SDTP, there exists $\Lambda \in F \cap \Omega \cap F_1(\Omega) \cap F_2(\Omega)$. Consequently, we have the following convex program for approximating the SDTP:

$$\begin{aligned} (\text{LP}(\Omega)) \quad \max_{\Lambda} \psi(\Lambda) &= \sum_{l=1}^{L} \sum_{i=1}^{S} [(c_i^l - m_i^l) x_i^l - \eta_i^l] \\ \text{s.t. } & \Lambda \in F \cap \Omega \cap F_1(\Omega) \cap F_2(\Omega) \ . \end{aligned}$$

To solve SDTP globally, we can implement a branch-and-bound algorithm which is proven to converge to a global optimal solution based on the above approximation scheme, where partitioning is performed by decomposing Ω into sub-hyper-rectangles. An outline of the algorithm is as follows:

- Initialization Step: The initial problem is the problem $LP(\Omega)$. Initialize $\Omega^{(1,1)} = \Omega$ and let $T_1 = \{(1, 1)\}$ be the index set of a single node at iteration one of the branch-and-bound tree. Let $UB^{(1,1)} = \infty$ be the upper bound associated with node $(1, 1)$. Let $LB = -\infty$ and $UB = \infty$ be the initial lower and upper bounds of the problem. Set $k = 1$, and go to the Main Step.

- Main Step: At iteration k, select a node (u, v) from T_k and remove this node from T_k. Solve LP($\Omega^{(u,v)}$) to obtain the partial solution $(\bar{x}, \bar{y}, \bar{z}, \bar{\pi})$. If $(\bar{x}, \bar{y}, \bar{z}, \bar{\pi})$ satisfies the constraint (3–9) and $\phi(\bar{x}, \bar{y}, \bar{z}, \bar{\pi}) = \psi(\bar{\Lambda})$, then the algorithm terminates with $(\bar{x}, \bar{y}, \bar{z}, \bar{\pi})$ as an optimal solution to SDTP(Ω). Otherwise, there are two possible cases: i) if $(\bar{x}, \bar{y}, \bar{z}, \bar{\pi})$ satisfies constraint (3–9) but $\phi(\bar{x}, \bar{y}, \bar{z}, \bar{\pi}) < \psi(\bar{\Lambda})$, then let

$LB_k = \phi(\bar{x}, \bar{y}, \bar{z}, \bar{\pi})$ be the current iteration lower bound, and set the partitioning index (p, q) as follows:

$$(p, q) = \arg\max_{(i,l)} \{\bar{z}_i^l \bar{x}_i^l - \bar{\eta}_i^l\};$$

else ii) if $(\bar{x}, \bar{y}, \bar{z}, \bar{\pi})$ does not satisfy constraint (3–9), then let $LB_k = -\infty$ be the current iteration lower bound, and set the partitioning index (p, q) as follows:

$$(p, q) = \arg\max_{(i,l)} \left\{ \max \left[\bar{c}_i^l x_i^l + \sum_{j \neq i} \bar{t}_{ij}^l y_{ij}^l - \bar{z}_i^l \bar{x}_i^l - \bar{\zeta}_i^l, - \sum_{\tau=l}^{L} (d_i^l \bar{\pi}_i^\tau + \alpha_i^l \gamma_i^l \bar{z}_i^l \bar{\pi}_i^\tau + \bar{\nu}_i^{l\tau}) \right] \right\} .$$

After finding (p, q), partition the region $\Omega^{(u,v)}$ into two mutually exclusive and exhaustive subregions. First, notice that $\Omega^{(u,v)}$ is a hyper-rectangle which can be expressed as follows:

$$\Omega^{(u,v)} \equiv \left\{ (x, y, z, \pi) \in \Omega : \ ZL_i^l \leq z_i^l \leq ZU_i^l \quad \forall i = 1, \dots, S, l = 1, \dots, L^P \ \right\} ,$$

where ZL_i^l and ZU_i^l are the lower and upper bounds of the component z_i^l. Using the above notation, the two new subregions can be represented as follows:

$$\Omega^{(k+1,1)} = \Omega^{(u,v)} \cap \{ZL_p^q \leq z_p^q \leq \bar{z}_p^q\}$$
$$\Omega^{(k+1,2)} = \Omega^{(u,v)} \cap \{\bar{z}_p^q \leq z_p^q \leq ZU_p^q\} .$$

Set $UB^{(k+1,1)} = UB^{(k+1,2)} = \psi(\bar{\Lambda})$. Then add these two nodes to the set T_k and update the tree, if necessary. Set $k = k + 1$, and go to the next iteration.

The detailed branch-and-bound algorithm and its updating operations are described in the Appendix D.

Al-Khayyal and Falk [9] showed that this algorithm converges to a globally optimal solution for a GBP; however, for our problem, the convergence rate for this algorithm can be disappointing, even for small-sized instances. In the next section, we therefore develop a specific penalty-based method for solving the SDTP.

3.3.3.3 Penalty-based method

We next discuss a customized penalty-based method for solving the SDTP by exploiting a property of constraint (3–9). Observe that the SDTP is a jointly constrained bilinear program without constraint (3–9), and jointly constrained bilinear programs can be solved using a suitable Reformulation-Linearization Technique (RLT).

As a result of the above observation, the first step of our method is to obtain a relaxation of the SDTP by eliminating constraint (3–9) from the constraint set and penalizing violations of this constraint in the objective function (3–8). Since the left-hand side of (3–9) is always greater than or equal to the right-hand side, penalizing violations of constraint (3–9) yields the following relaxed problem:

$$\max_{x,y,z,\pi} \sum_{l=1}^{L} \sum_{i=1}^{S} \left(c_i^l - m_i^l - z_i^l\right) x_i^l - M \left\{ \sum_{l=1}^{L} \sum_{i=1}^{S} \left[\left(\bar{c}_i^l - z_i^l\right) x_i^l + \sum_{j \neq i} \bar{t}_{ij}^l y_{ij}^l - \sum_{\tau=l}^{L} \pi_i^\tau \left(d_i^l + \alpha_i^l \gamma_i^l z_i^l\right) \right] \right\}$$

$$\text{s.t.} (x, y, z, \pi) \in F \cap \Omega \ .$$

where M is a sufficiently large positive number (which corresponds to a penalty per unit of violation of the constraint). By rewriting the objective, we obtain the following equivalent formulation:

$$\text{(PEN)} \max_{x,y,z,\pi} \sum_{l=1}^{L} \sum_{i=1}^{S} \left\{ (c_i^l - m_i^l - M\bar{c}_i^l) x_i^l - M \sum_{j \neq i} \bar{t}_{ij}^l y_{ij}^l + M \sum_{\tau=1}^{l} d_i^\tau \pi_i^l + (M-1) z_i^l x_i^l + M \sum_{\tau=l}^{L} \alpha_i^l \gamma_i^l \pi_i^\tau z_i^l \right\}$$

$$\text{s.t.} (x, y, z, \pi) \in F \cap \Omega \ .$$

The above formulation is a jointly constrained bilinear program. Sherali and Alameddine [84] developed an RLT for this problem class and embedded it within a provably convergent branch-and-bound algorithm. Sherali and Alameddine's RLT reformulates the bilinear program by first constructing valid nonlinear inequalities from the original constraints defining $F \cap \Omega$. The following are two general methods for generating these additional nonlinear inequalities:

- Multiplying any two constraints in Ω pairwise, e.g., $(c_i^l - m_i^l - z_i^l)(U^k - x_j^k) \geq 0$; and
- Multiplying a bounding constraint in Ω with a constraint in F, e.g., $(c_i^l - m_i^l - z_i^l)(\sum_{\tau=k}^{L} \pi_j^\tau + z_j^k - \bar{c}_j^k) \geq 0$.

Defining the set of constraints generated using the above pairwise product operations as $F(\Omega)$, then the original *PEN* problem constraints together with the constraints in $F(\Omega)$

yield a new equivalent formulation of the problem PEN, which we denote as PEN':

$$\text{(PEN')} \max_{x,y,z,\pi} \sum_{l=1}^{L}\sum_{i=1}^{S}\Bigg\{ \left(c_i^l - m_i^l - M\bar{c}_i^l\right) x_i^l - M\sum_{j\neq i}\bar{t}_{ij}^l y_{ij}^l + M\sum_{\tau=1}^{l} d_i^\tau \pi_i^l + (M-1) z_i^l x_i^l + M\sum_{\tau=l}^{L}\alpha_i^l\gamma_i^l\pi_i^\tau z_i^l \Bigg\}$$

$$\text{s.t. } (x,y,z,\pi) \in F \cap \Omega \cap F(\Omega) \ .$$

All of the nonlinear terms of the PEN' formulation are bilinear and, as a result, PEN' can be linearized through an appropriate variable substitution strategy, which transforms the nonlinear constraints of the set $F(\Omega)$ to a set of linear constraints. For example, we substitute:

$$\eta_i^{kl} = x_i^k z_i^l \forall i = 1, \dots, S, k, l = 1, \dots, L \ ,$$

$$\nu_i^{kl} = \pi_i^k z_i^l \forall i = 1, \dots, S, k, l = 1, \dots, L \ .$$

Let ζ represent the vector containing all such new variables created other than η and ν, and let $F_l(\Omega)$ represent the linearized set of constraints from $F(\Omega)$. This leads to the following reformulation of PEN':

$$\max \sum_{l=1}^{L}\sum_{i=1}^{S}\Bigg\{ \left(c_i^l - m_i^l - M\bar{c}_i^l\right) x_i^l - M\sum_{j\neq i}\bar{t}_{ij}^l y_{ij}^l + M\sum_{\tau=1}^{l} d_i^\tau \pi_i^l + (M-1)\eta_i^{ll} + M\sum_{\tau=l}^{L}\alpha_i^l\gamma_i^l\nu_i^{l\tau} \Bigg\}$$

$$\text{s.t. } (x,y,z,\pi,\nu,\eta,\zeta) \in F \cap F_l(\Omega) \cap \Omega \ .$$

Note that after linearization, the resulting problem is a relaxation of the PEN' problem, which corresponds to an upper bounding linear program for the original bilinear program; Sherali and Alameddine [84] showed that the resulting upper bound is at least as good as that obtained using Al-Khayyal's approach. The resulting branch-and-bound algorithm we use is very similar to Algorithm 2 of Al-Khayyal's approach. There is only one difference: the branch-and-bound algorithm in our method does not need to check whether the partial solution $(\bar{x}, \bar{y}, \bar{z}, \bar{\pi})$ is feasible, so the partitioning index (p, q) can be

found as follows:

$$(p, q) = \arg\max_{(i,l)} \left\{ \max \left[\nu_i^{ll} - z_i^l x_k^l, \sum_{\tau=l}^{L} \alpha_i^l \gamma_i^l (\nu_i^{l\tau} - \pi_i^\tau z_i^l) \right] \right\} \quad .$$

The remainder of our branch-and-bound algorithm uses the same updating operations and stopping criteria as Algorithm 2 from Al-Khayyal [7].

From the above description, we derive a procedure for solving the SDTP using the penalty-based method and RLT as shown in Algorithm 1.

Algorithm 1 Penalty-based approach

$M \leftarrow 0$, $UB \leftarrow \infty$ and $LB \leftarrow -\infty$;
while $UB - LB > \epsilon$ **do**
  $(x', y', z', \pi') \leftarrow$ an optimal solution corresponding to PEN';
  $UB \leftarrow$ the optimal objective value of PEN';
  $(\bar{x}, \bar{y}, \bar{\pi}) \leftarrow$ an optimal solution corresponding to SDTP problem for fixed vectors z';
  $LB_k \leftarrow \sum_{l=1}^{L} \sum_{i=1}^{S} \left[c_i^l - m_i^l - (z')_i^l \right] \bar{x}_i^l$;
  if $LB < LB_k$ **then**
    $LB \leftarrow LB_k$ and $(x^*, y^*, z^*, \pi^*) \leftarrow (\bar{x}, \bar{y}, z', \bar{\pi})$;
  end if
  $M \leftarrow M + \beta$;
end while.

In Algorithm 1, β is a positive number. At a given iteration, our approach first uses the branch-and-bound algorithm to obtain an exact optimal solution or near optimal solution to PEN', and then updates the best feasible solution and the lower and upper bounds of the SDTP problem. If the optimality gap is less than a predetermined tolerance, then the algorithm terminates and returns the current best feasible solution; otherwise, the algorithm increases the value of penalty term M by a constant β.
From our numerical tests, we found that for smaller values of M, the associated branch-and-bound algorithm converges faster and has a greater chance of finding a good feasible solution earlier; for a large enough value of M, the PEN' problem gives the same optimal solution as SDTP, and this value of M was usually under 10 for our test instances.

3.4 Numerical Experiments

This section presents computational test results for our penalty-based approach, the iterative LP heuristic approach, and Al-Khayyal's approach for solving the SDTP. We will demonstrate that our penalty-based approach outperforms the other two solution methods from the literature. In addition to evaluating the performance of different solution approaches, we also analyze the impacts of different parameters on the bullwhip effect and its associated costs, as well as on the supplier's net profit from wholesale discounts. We implemented all three of the solution approaches in the C# programming language, with the relaxed linear programs solved using ILOG®'s CPLEX® 12.5 solver with Concert Technology. We performed all tests on a computer with an Intel® Dual Core 1.70 GHz and 6 GB memory.

3.4.1 Comparison of Solution Methods

To benchmark the performance of our penalty-based approach with the iterative LP heuristic and Al-Khayyal's approach, we tested our solution method using nine problem sets. Each problem set corresponds to a fixed number of retail stores and number of time periods, (S, L), where $S \in \{2, 3, 4\}$ and $L \in \{2, 3, 4\}$. We tested twenty randomly generated instances for each combination of (S, L) values, for a total of 180 problem instances. For both exact solution approaches, we set the relative optimality tolerance to 10^{-2}, and the time limit to 180 seconds.

Table 3-1 summarizes the distributions used in generating parameters in our computational study. For each problem instance, the supplier's unit wholesale prices c_i^l, deterministic components of consumer demand d_i^l, and retailer's pass through rates α_i^l were generated from uniform distributions first. Then, based on the generated values of c_i^l and d_i^l, the supplier's unit costs m_i^l, the retailer's unit inventory holding costs h_i^l, the retailer's unit transportation costs t_{ij}^l and the consumer promotional price elasticity γ_i^l were generated based on continuous uniform distributions. We let $U(l, u)$ denote the continuous uniform distribution with lower bound l and upper bound u.

Table 3-1. Parameter distributions used in computational tests.

Supplier's unit wholesale price, c_i^l $U(10.00, 20.00)$	Supplier's unit cost, m_i^l $c_i^l \times U(0.50, 0.80)$
Retailer's unit inventory holding cost h_i^l $c_i^l \times U(0.01, 0.20)$	Retailer's unit transportation cost, t_{ij}^l $\lvert c_i^l - c_j^l \rvert \times U(1.01, 1.20)$
Deterministic component of consumer demand, d_i^l $U(100, 500)$	Retailer's pass-through rate, α_i^l $U(0.30, 1.00)$
Consumer promotional price elasticity, γ_i^l $\frac{d_i^l \times c_i^l}{20} \times U(0.10, 0.50)$	

Observe that the largest size instance in our computational study only considers four stores, and in practice, it is not common that a company operates a small number of retail stores but still serves geographically dispersed areas. However, retailers like Wal-Mart® usually use a spoke-and-hub strategy: they first move the products from suppliers to distribution centers, and then from distribution centers to local stores. By adopting this strategy, retailers only need to operate a small number of distribution centers. In fact, at the start of 2010, 40 out of the top 75 food retailers in North America had no more than four distribution centers in U.S. and Canada [75]. The distribution centers are usually geographically dispersed across the country, and the local store orders will be aggregated at the distribution center. Moreover, retail stores served by the same distribution center will have the same wholesale price, and as a result, transshipment will only occur between the distribution centers. Our choice of parameters match well with this type of spoke-and-hub strategy, and the only change that needs to be made is in using distribution centers instead of stores in our model. In addition, the reason we selected the number of planning periods to be no more than four is because we assume the retailer can make accurate demand forecasts over the planning horizon; as a result the number of periods in the planning horizon cannot be too large, assuming that each time period represents one or two weeks, for example.

For each problem instance, in addition to the price promotion game, we also consider a "no-discount" case and a basic promotion case. In the "no-discount" case, the supplier does not offer any price discount to the retailer over the entire planning horizon; in this case the retailer's optimal profit is PR_0 and the supplier's optimal profit is PM_0. In the basic promotion case, the supplier sets its discount policy by assuming the retailer will pass the entire discount on to its consumers and will neither forward buy nor divert (when, in fact, the retailer will minimize its cost by applying both forward buying and diverting strategies). For both promotion cases, we define the retailer's optimal profit and the supplier's profit as PR_1 and PM_1, respectively. The performance measures we will use for comparative purposes include the bullwhip effect, $BWE = \sqrt{\frac{Var[x]}{Var[d]}}$, the retailer's profit gain, $\Delta_r = \frac{PR_1 - PR_0}{PR_0} \times 100\%$, the supplier's profit gain, $\Delta_m = \frac{PM_1 - PM_0}{PS_0} \times 100\%$, and the total system profit gain, $\Delta = \frac{PS_1 + PR_1 - PS_0 - PR_0}{PS_0 + PR_0} \times 100\%$.

To compare our algorithm with Al-Khayyal's approach, we consider the running time and relative optimality performance. The results of our tests, averaged over the twenty random problem instances for each combination of (S, L) values, are presented in Table 3-2. The table shows that for smaller-size problems with $(S, L) \in \{(2,2), (2,3), (2,4), (3,2), (3,3), (4,2)\}$, our approach on average takes less than 150 seconds to reduce the relative optimality gap under 3%. For the same set of the problems, except for problems with $(S, L) = (2,2)$, the Al-Khayyal's approach could not solve the problems globally within the 180 second time limit, and the average relative optimality gap is unacceptably large (72.55%). For large-size problems with $(S, L) \in \{(3,4), (4,3), (4,4)\}$, neither approach was able to reduce the relative optimality tolerance under 5% within 180 seconds. However, the solutions given by our approach have much smaller relative optimality gaps when compared with the solutions obtained using Al-Khayyal's approach.

We also tested each problem instance using the commercial nonlinear programming solvers GAMS/LINDOGlobal and GAMS/BARON. Both of these solvers only guarantee

Table 3-2. Computational results I.

Problem set		Al-Khayyal		Penalty method	
S	L	Relative gap (%)	Time	Relative gap (%)	time
2	2	4.46	147.54	0.91	28.52
	3	49.53	180.00	1.57	59.85
	4	78.36	180.00	2.33	123.74
3	2	17.02	180.00	1.34	54.33
	3	114.41	180.00	2.58	141.48
	4	126.61	180.00	5.29	180.00
4	2	103.45	180.00	2.38	125.05
	3	126.89	180.00	5.35	180.00
	4	134.02	180.00	8.78	180.00
5	2	119.21	180.00	2.94	130.91

local optimality of their solutions, and they failed to obtain meaningful upper bounds for the problem with $(S, L) \in \{(3,4),(4,3),(4,4)\}$ for all instances. For the other problems with a meaningful upper bound, the average relative gap is more than 4% and the maximum relative gap is 9.3%.

In addition to relative optimality gaps, we also consider the best feasible solution obtained as another performance criterion. As defined above, PM_0 is the supplier's net profit for the case of "no discount," and PM_1 is the supplier's net profit for the promotion cases. The quantity $\Delta_m = \frac{PM_1 - PM_0}{PS_0} \times 100\%$ measures the performance of the promotion case compared with the "no discount" case. If $\Delta_m > 0$, this means the corresponding promotion case is more profitable than the "no discount" case; otherwise, the "no discount" case is a better option for the supplier. Moreover, a larger value of Δ_m means a greater level of profit for the corresponding promotion plan. From Table 3-3, we observe that our approach found better feasible solutions than Al-Khayyal's approach on average, and the general promotion game case outperforms the basic promotion case and the case of "no discount" on average. However, the basic promotion case is not necessarily more profitable than the "no discount" case. Another interesting observation is that the best feasible solution obtained from our approach has a smaller value of BWE than in the basic promotion case on average, which is consistent with

Table 3-3. Computational results II.

Problem instance		Base		Al-Khayyal	Penalty method	
S	L	$\Delta_m(\%)$	BWE	$\Delta_m(\%)$	$\Delta_m(\%)$	BWE
2	2	-11.01	8.01	3.12	10.10	4.35
	3	-5.58	6.85	-7.94	5.92	5.16
	4	-8.26	6.32	-14.25	4.23	3.40
3	2	-8.10	5.48	2.04	5.05	4.92
	3	-8.12	7.87	-19.69	4.29	5.91
	4	-8.99	7.33	-15.26	3.98	4.83
4	2	-8.82	4.61	-20.63	3.12	3.99
	3	-5.83	6.46	-20.15	3.30	4.24
	4	-6.97	6.97	-19.01	3.03	5.25

the findings of Lee, Padmanabhan, and Whang [64] that the bullwhip effect impairs upstream performance. On the other hand, the BWE on average is greater than one for the best solution obtained from our approach, which indicates that the revenue gain from price promotions can compensate for the extra cost induced by the bullwhip effect if the supplier takes the retailer's reactions into consideration and judiciously applies a trade promotion strategy.

To test the performance of the successive linear programming heuristic, we set the initial value of the promotion vector z to 0. We continued to use 180 seconds as the time limit. If the improvement between two successive iterations is less than 10^{-2}, we stop the heuristic even if it is before reaching the time limit. The results derived from the heuristic are very disappointing: even when there is an obvious solution better than the "no discount" case (for example, the basic promotion case usually gives a better solution than "no discount case"), the heuristic always stops at $z = 0$ after two iterations. The reason for this is as follows: assuming that (x^0, y^0, π^0) optimizes the objective function for the initial value of z, then the value of z that optimizes the objective function for the fixed (x^0, y^0, π^0) is still a vector of zeroes because, as we can observe from the objective function, for a fixed value of x, it is always optimal to set z to 0 when this is feasible.

3.4.2 Parameter Analysis

The goal of this section is to study how the pass-through rate, α, and consumer promotional price elasticity, γ, influence the profit performance $(\Delta, \Delta_m, \Delta_r)$ and the BWE for the case of $(S, L) = (3, 3)$. To this end, we considered ten levels of α and three levels of γ when applied to three randomly generated problem instances, for a total of 90 additional test cases. In this section, we consider the results obtained by solving each of these test problem instances using our penalty-based approach, because of its ability to consistently obtain a superior feasible solution for SDTP for the previously tested instances.

Figures 3-1-3-4 illustrate the results of these experiments. The results shown in the figures lead to the following observations.

Figures 3-1-3-3 show that, for all problem instances, a higher value of γ implies higher values of Δ_m, Δ_r and Δ. This effect is quite intuitive, because when consumers are more responsive to price reductions, we expect that the performance of both the supplier and the retailer will improve as a result.

Figure 3-1 shows that, for all problem instances, a larger value of the pass-through rate, α implies larger values of Δ_m. This observation is consistent with the fact that a low pass-through rate for trade promotions is a major cause of inefficiency in trade promotions.

In Figure 3-2, for problem instances 1 and 2, the retailer's profit gain decreases as the pass-through rate increases when the level of γ is low. So, when the retail price cut does not attract a sufficient number of additional consumers, the retailer has incentive to lower the pass-through rate to gain more profit. However, when the level of γ is high, the retailer's optimal pass-through rates are usually non-zero, and sometimes the best decision for the retailer is to pass through more than 100% of the supplier's trade promotion to consumers, as shown for instance 3.

Figure 3-3 shows that when the value of γ is sufficiently large, Δ may initially decrease in α (or remain unaffected), and then, at some positive value of α, begin increasing in α. However, when the level of γ is low, Δ always decreases in α for instances 1 and 2. Thus, from a system perspective, the supplier's choice of whether or not to offer a discount depends on the value of γ.

Figure 3-4 shows an interesting pattern with respect to the bullwhip effect. For all instances, the bullwhip effect may initially increase or remain unchanged in α; it then jumps to a much higher value at some positive value of α, after which it ultimately decreases in α. We observe that the "jump" in the bullwhip effect occurs at the same point at which the increase in Δ_r suddenly becomes large. We also find that the larger the value of γ, the earlier this "jump" occurs. From our numerical tests, we found that there are typically two types of promotion plan patterns for each problem instance, which we will call "Plan L" and "Plan H," respectively. Price discounts for "Plan H" are much deeper than price discounts for "Plan L". However, when the pass-through rate α is low, it is optimal for the supplier to adopt "Plan L" instead of "Plan H". As the value of α increases, after reaching a certain positive value of α, the optimal promotion plan for the supplier changes from the form of "Plan L" to "Plan H," and we call this the threshold value of α. When the value of γ is large, this means the consumers are more sensitive to retail price discounts, so, for the same level of retail discount, a higher value of γ means more new consumers, which results in a smaller threshold value of α. Because the number of new consumers attracted by "Plan H" is usually much larger than that of "Plan L," "Plan H" typically brings a larger system profit and a higher value of the bullwhip effect.

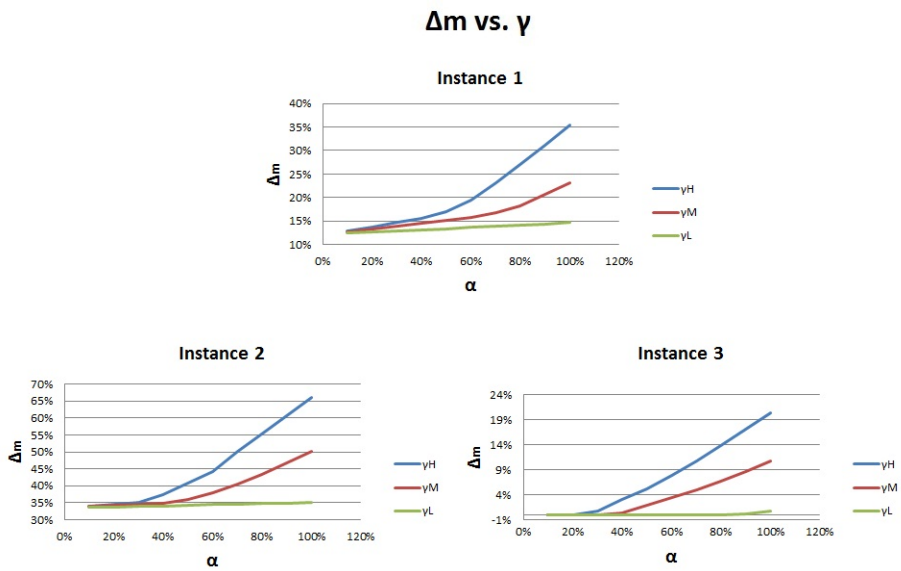

Figure 3-1. Supplier profit gain in α with γ at different levels.

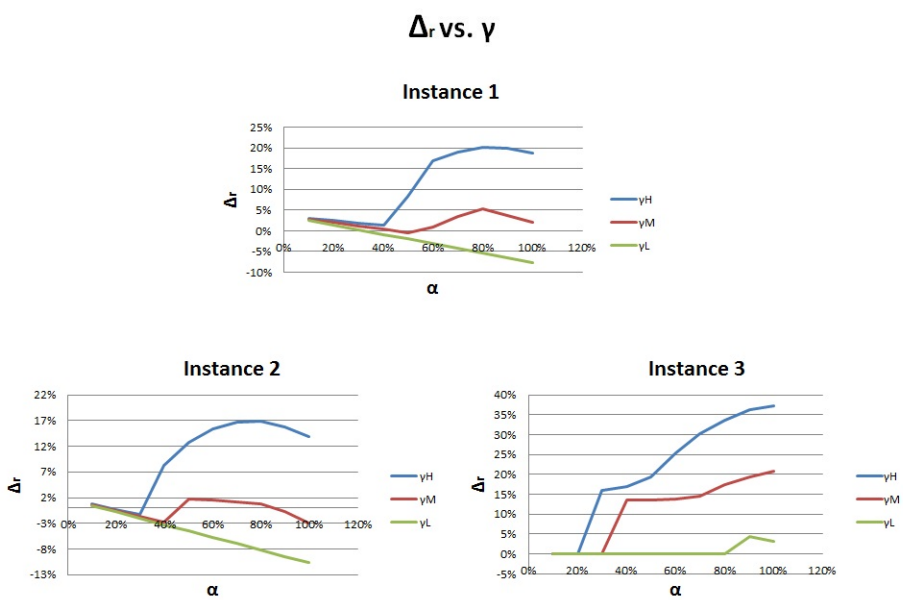

Figure 3-2. Retailer profit gain in α with γ at different levels.

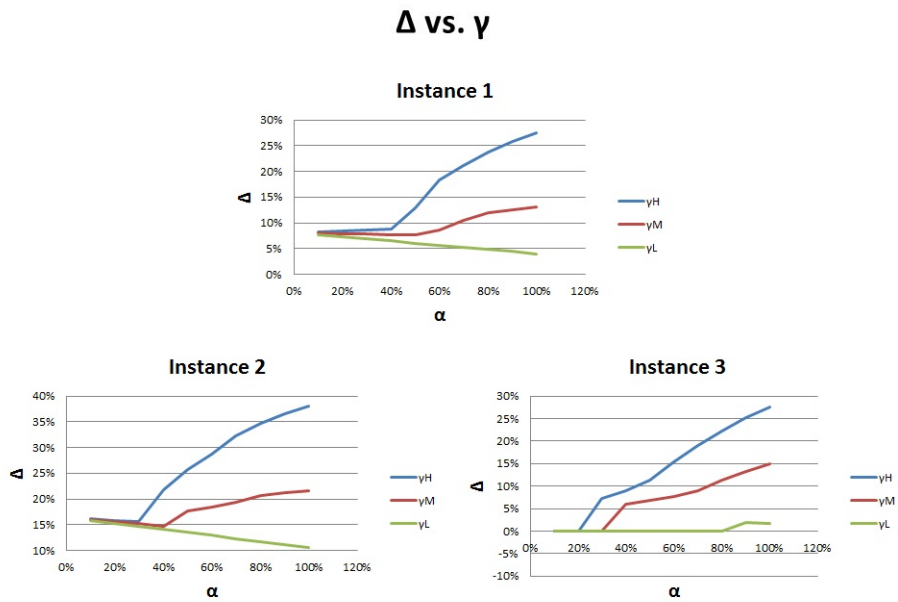

Figure 3-3. Total system profit gain in α with γ at different levels.

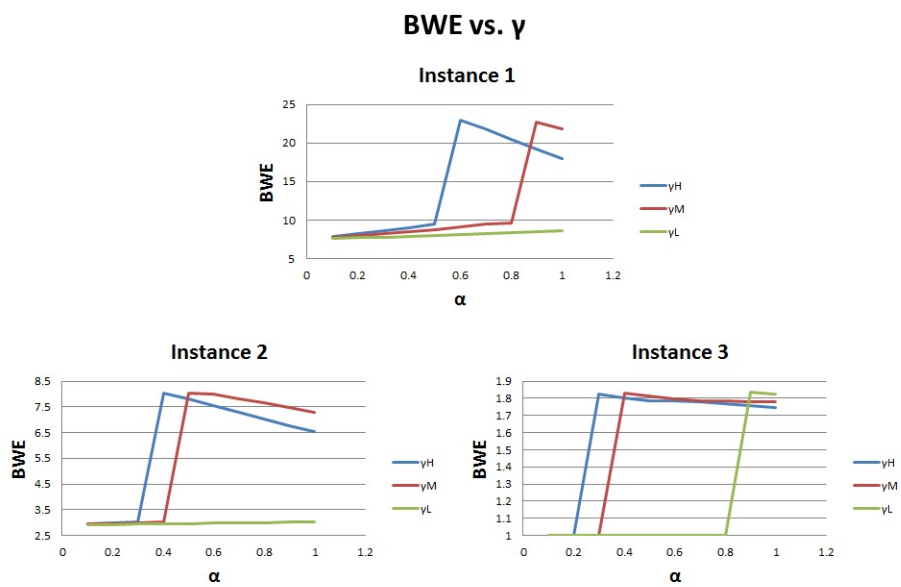

Figure 3-4. Bullwhip effect in α with γ at different levels.

CHAPTER 4
SALES PROMOTIONS AND ASSORTMENT PLANNING IN MULTIPLE SELLING CHANNELS

4.1 Problem Description and Literature Survey

U.S. B2C e-commerce has been booming in recent years, and the total sales grew from 72 billion U.S. dollars in 2002 to 256 billion U.S. dollars in 2011 [63]. In this prosperous time of online shopping, a so called "clicks-and-mortar" business model has emerged. In this business model, the online retail channel is integrated with the traditional retail channel. Both traditional retailers and previous pure-play Internet retailers have contributed to the development of the clicks-and-mortar business model. On the one hand, traditional retailers like Wal-Mart® and Target® have to limit their assortment only to popular items at the stores because of the high number of products competing for the limited store space [43]. However, Mittelstaedt and Stassen [73] showed that such reductions in assortment encourage "variety-seeking" shoppers to visit other stores, which leads to potential retailer financial loss. To overcome space limitations, many of the largest retailers (such as Wal-Mart®, Target® and Best Buy®) have adopted the clicks-and-mortar model to make a broader assortment available at their online stores. On the other hand, pure-Internet retailers are also moving into offline shopping by opening physical stores or collaborating with traditional retailers [67], as in the case of Warby Parker®, a successful online eyeglass retailer, which just opened its first physical store in New York City in 2013. One of the reasons for the move by these online retailers is that certain categories of products are less amendable to online shopping, so an online-only retail channel is not enough to ensure long-term success. For instance, it is hard for a pure-Internet fashion retailer to succeed in the long term, because consumers need to try and feel the product beforehand. Lack of physical experience in real time causes a high return rate (up to 45%) associated with the online apparel sales [2], and costs of return handling usually are high for an online-only retailer. In addition, the competition from traditional retailers like Wal-Mart® who are actively

expanding their visual channel also urges the pure-Internet retailers to move into the offline-shopping world.

In this study, a decision tool that helps retailers coordinate product assortments, retail prices and inventory between the physical channel and online channel is developed in the hope to reduce channel conflicts and improve the retailer's profitability. The research in coordinating traditional and Internet channels is voluminous. Agatz et al. [2] provided an excellent review which addresses various supply chain management issues specific to e-fulfillment in a clicks-and-mortar model. In their review, Agatz et al. discussed both practical managerial planning issues and corresponding operations research models. According to their review, management issues that are related to our study include pricing, distribution network design and inventory and capacity management.

First, pricing decisions play an important in role in any business model, and a significant stream of research is available for multi-product pricing optimization, which appears to be applicable to a clicks-and-mortar business model. The Multinomial Logit (MNL) model and the Nested Logit (NL) model are mainly used as models of consumer choice. Studies by Hanson and Martin [56], Dong et al. [41], Song and Xue [89] and Chen and Hausman [37] are examples of recent works which used the MNL model to simultaneously determine prices for substitutable products. Although the MNL model is widely used in the multi-product pricing literature, the independence of irrelevant alternatives (IIA) property of the MNL model restricts its application. The IIA property generally does not hold in the case where the consideration set could be divided into subsets such that products within one subset are more similar to each other than across subsets. To address this drawback, Li and Huh [66] and Gallego and Wang [49] considered the multi-product pricing problem with the NL model.

Distribution network design is another key strategic decision in a clicks-and-mortar supply chain. With the development of the technique called drop-shipping[1] , clicks-and-mortar retailers (e.g., Staples® Inc.) have started adopting a dual supply chain strategy. They still hold their own inventory for high sales volume products, but they deliver other less popular products using drop-shipping [86]. In this way, they can achieve benefits in terms of both wider product selection and lower inventory costs. Randall et al. [82, 83] discussed the costs and benefits of using the drop-shipping technique and the trade-offs between traditional and virtual supply chain structures. Bailey and Rabinovich [18] developed an analytical model which examines an online book retailer's decisions under in-stock and drop-shipping inventory strategies, and they showed that the retailer should adopt both inventory management strategies as products become more popular and the retailer increases its market share. Netessine and Rudi [78] developed a noncooperative game to model the dual inventory strategy (i.e., in-stock and drop-shipping strategies) of a two-level supply chain with a single wholesaler and multiple retailers, and their study provides practicing managers with guidelines for choosing appropriate distribution channels.

The third management issue related to our study is inventory and capacity management. Inventory management in the clicks-and-mortar business model can be very flexible. For example, the clicks-and-mortar retailer could make in-store inventories available to online buyers or deliver out-of-stock items directly to local consumers by using virtual channel inventory. The clicks-and-mortar inventory management issues usually arise from the interactions between different consumer segments which can be addressed by inventory rationing models. Cattani and Souza [33], Ayanso et al. [16] and

[1] The drop-shopping technique is an outsourcing process in which the retailer transfers the consumer order and shipment details to the manufacturer or to a wholesaler, who then ships the order directly to the consumer at the cost of a higher wholesale price.

Ding et al. [40] all made contributions to the literature in the area of inventory rationing in the online channel.

In addition to the three issues mentioned in [2], another management issue related to our study is assortment planing. Academic study on assortment planning is relatively new, but it has experienced quick growth recently. Kök et al. [61] did an extensive review of the assortment planning literature. An early model for assortment planning is proposed by van Ryzin and Mahajan [95]. In this model, van Ryzin and Mahajan used a MNL to describe the consumer choice process and a newsvendor model to describe the supply process. They assumed that the set of possible variants is homogeneous, which means each variant is sold at an identical price and charged an identical cost. They showed that, under these assumptions, the optimal assortment can be restricted to one of n possible types, where n is the number of number of variants under consideration. The result of this study is elegant, but few practical problems fit the strict assumptions of this model. In their second paper, Mahajan and van Ryzin [70] still studied a newsvendor-like model with consumer choice decisions based on MNL. But in this study products can have nonidentical price and cost, and consumers dynamically substitute among products when a stock-out occurs. Their results from this study showed that the stock level of a popular variant should be higher than the stock level suggested by a traditional newsvendor analysis under dynamic substitution. Cachon et al. [32] extended the van Ryzin and Mahajan [95] model by incorporating consumer search, and they showed that in the presence of consumer search it may even be optimal to add a nonprofitable product to an assortment so as to prevent consumer search. Another important assortment planning study based on the MNL model was developed by Miller et al. [71]. Their study used consideration sets to capture the heterogeneity of consumer preference and used integer programming to determine the optimal retail assortments.

In addition to the MNL model, the exogenous demand (EXD) model is another commonly used model in the assortment planning literature. The EXD model overcomes the shortcomings of MNL model and has more degrees of freedom. The EXD model directly specifies the demand for each product and the substitution rate if the product is not available, but there is no underlying utility model to explain the consumer behavior. Smith and Agrawal [87] considered the problem of optimizing assortments with the EXD model. In their model, they assumed that the consumer purchase behavior is dynamic, which means a consumer's final choice depends on the consumer's preferences, the availability of all products upon the arrival of the consumer, the choice of previous consumers, and the number of substitution attempts made by the consumers. Their optimization model is computationally infeasible in most cases, so they developed an approximation method to solve it.

One group of problems that is strongly relevant to the assortment planning problem is shelf space allocation planning problems. Under the assumption that allocating zero space to a product is equivalent to eliminating it from the assortment, shelf space allocation and assortment selection can be determined simultaneously. In contrast to the assortment planning literature which assumes shelf space has no effect on consumer preferences, nearly all the shelf space allocation literature is based on the assumption that space allocated to products has effects on sales, and these effects are measured by space elasticity. The shelf space allocation problem has been considered by Anderson and Amato [10], Hansen and Heinsbroek [55], Corstjens and Doyle [39], Borin and Farris [27], Urban [94], Hwang et al. [58] and many other researchers. For the literature on shelf space allocation, an often-stated criticism is that the relationship between sales and space allocation is weak, and the impact of shelf space allocation on sales is very small relative to other marketing variables (e.g., advertising and promotion) [101].

4.2 Problem Statement and Formulation

We consider a clicks-and-mortar retailer who sells goods to end-consumers through both physical and virtual (e-commerce) channels. This retailer operates two physical facilities (a warehouse and a retail store) and a website for online shopping. Goods can be stored in the warehouse and the store, and may be transferred from both facilities to consumers in direct shipments. The retailer has limited storage capacity, and would like to determine the set of products to offer, both in the store and online, in order to maximize its profit from sales during a single selling season.

4.2.1 Product Related Assumptions

Let $N_1 = \{1, \dots, n\}$ and $N_2 = \{n+1, \dots, 2n\}$, respectively, denote the product selection sets at the local and online stores. We assume product $i \in N_1$ and product $i + n \in N_2$ are identical, except that i is sold at the local store but $i + n$ is sold at the online store. We make this assumption for the sake of notational simplicity, and because it also reflects the fact that product i is different from product $i + n$ from the consumer's perspective and should be treated differently when kept in stock during the selling season. We thus make the assumption that, during the selling season, the inventory of product i can only be used to satisfy demands for product i; at the end of the season, however, any leftover inventory of product $i \in N_1$ may be used to satisfy outstanding orders for product $i + n \in N_2$. We will discuss the specifics of order fulfillment later in greater detail.

In addition, each of the products is characterized by P possible selling prices. The upper limit for the price of product i, which is also the highest possible selling price, r_{i1}, is the original retail price of product i before any discounting. The lower limit of the price of product i is the wholesale price of product i, which is denoted as c_i. We define the pth highest possible selling price as

$$r_{ip} = r_{i1} - \frac{(r_{i1} - c_i)(p-1)}{P}.$$

The value of P depends on the desired resolution. Moreover, for each possible selling price of the product i, we create a pseudo-product (i, p).

4.2.2 Facilities Related Assumptions

Without loss of generality, we assume that the local store has both displayed inventory and backroom inventory. Part of any received order is delivered to the shelves directly, with the remainder placed in backroom storage before being brought to the display area. The shelf space at the local store has three dimensions, namely width, depth and height as shown in Figure 4-1. In this study, we only consider the width dimension when allocating shelf space. The shelf width is divided into *facings*, and the size of product i's *facing* is equal to the width α_i of its front face, in other words, each individual unit of a product is one facing wide. Capacity of product i's facing depends on the height and depth of the shelf and the physical size of a unit of product i. The number of facings allocated to each product determines the maximum level of its displayed inventory. For example, there are two types of products (i.e., products i and j) in Figure 4-1. Two facings are allocated to product i, while four facings are allocated to product j. By observation, six units can be held in one facing of product i, so the total space allocated to product i can store 12 units. Similarly, the four facings allocated to product j can hold 36 units.

We further assume that the order quantity for each product is large enough to fill its initially allocated displayed inventory space (full stocked), and items are restocked from the backroom onto shelves as displayed items are depleted by consumer demands.

There are many possible ways for the clicks-and-mortar retailer to manage inventory at the local store backroom and the warehouse, and our model is sufficiently flexible to incorporate several types of inventory strategies. In order to concisely present our modeling methods, we assume that the distance between the warehouse and local store is too long to transship any product from one facility to the other during the selling

season. We also assume that all of the products in the set N_1 can only be stored at the local store, while the variants in N_2 can only be stored in the warehouse.

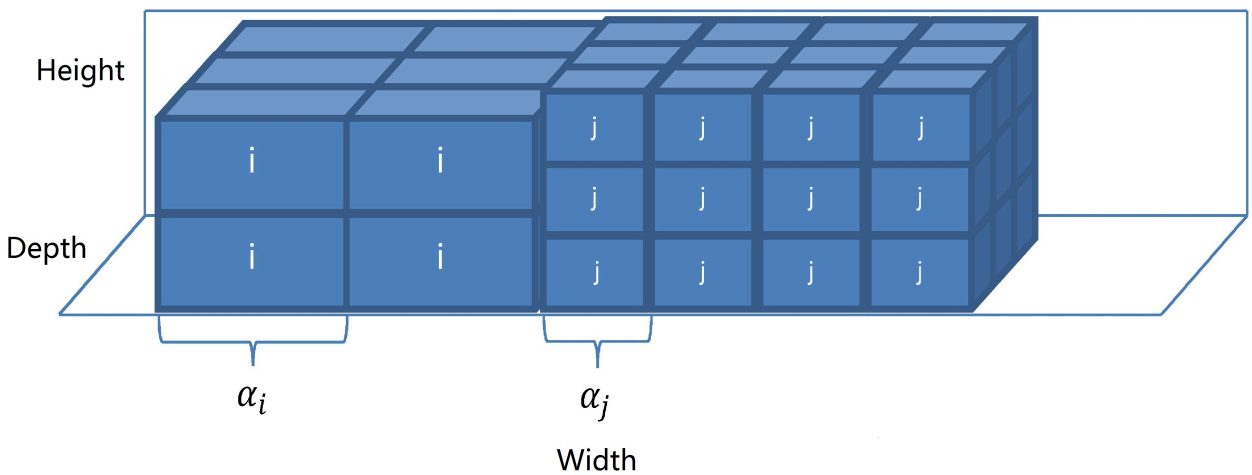

Figure 4-1. Illustration of shelf space allocation.

4.2.3 Demand Related Assumptions

Between the local and online stores, the retailer serves K consumer segments, and every consumer has a choice of purchasing either at the store or placing an order on the website. The segmentation of consumers into distinct groups may be based on consumer purchase behavior, consumer needs, channel preferences, and/or interest in certain product features. In our study, there are at least three consumer segments, namely walk-in, online and hybrid consumers. While walk-in consumers are assumed to only shop at the physical store, online consumers place orders on the website and have the order shipped directly to their residence. Moreover, hybrid consumers have a primary channel preference and are likely to switch to the other channel if their preferred product is not available in the preferred channel. These three basic segments can be

further divided into sub-segments to convey more useful information about consumer attributes.

Let $N \equiv N_1 \cup N_2$ denote the entire set of $2n$ products. One of the retailer's decisions is the set(s) of products to make available to consumers during the selling season. We call the subset $S \subseteq N$ of available products, the retailer's "offer set." The other retailer decision is to determine the selling price of each product selected in the offer set. Notice that if product i is selected in the offer set, one and only one of the P possible selling prices should be selected. The final outcome of these two retailer decisions is a set of pseudo-products, which is denoted as $SP \subseteq N \times \{1, \dots, P\}$. Given a pseudo offer set SP, an arriving consumer from segment k chooses product $(i, p) \in SP$ with probability $P_{kip}(SP)$, where $P_{kip}(SP) = 0$, if $(i, p) \notin SP$. We denote the no-purchase probability by $P_{k0}(S)$, and by total probability, we have that $\sum_{(i,p)\in SP} P_{kip}(SP) + P_{k0} = 1$. If the number of segment-k consumers arriving during the planning horizon follows a Poisson distribution with mean λ_k, and the choice of each consumer is independent of the others, then the demand of segment-k consumers for pseudo-product (i, p) when the retailer offers set SP follows a Poisson distribution with mean $\lambda_k P_{kip}(SP)$.

Before proceeding into details about the sequence of events, we make the following three important assumptions concerning the consumer choice process, which are very similar to the static substitution assumptions used by Smith and Agrawal [87], van Ryzin and Mahajan [95], Topaloglu [93].

Assumption 1. *Consumers choose based only on knowledge of the set SP, and their choices will be not be affected by the inventory levels and shelf space allocation of the pseudo-products in SP.*

Assumption 2. *If an in-store consumer selects a product in $S \cap N_1$ and the store does not have it in stock, the consumer does not undertake a second choice, and the sale is lost.*

Assumption 3. *If an online consumer selects a product in $S \cap N_2$ and the online store does not have it in stock, the consumer does not know this, and the sale is not lost. The retailer satisfies such a demand through either a product transfer between facilities or drop-shipping.*

Under our assumptions, for the traditional channel, a consumer's initial choice is only influenced by the set of alternatives offered, and there is no dynamic substitution if their first choice is out of stock. Van Ryzin and Mahajan [95] showed several examples where these assumptions serve as a reasonable approximation of consumer behavior in traditional selling channels. However, the consequences of stockouts in traditional channels could be dire: local store consumers may switch stores and purchase the item elsewhere, and this store switching could be permanent. As a result, it is both theoretically and practically desirable to dampen the negative effects of stockouts. To achieve this goal, we apply a chance constraint to limit the probability of stocking out at the local store. For online retail channels, our assumptions approximate the practice in which the stock level of product is not shown on the website, but the consumer order is always satisfied.

4.2.4 Sequence of events

The single selling season in our model can be divided into three stages: before the selling season, during the selling season, and after the selling season. We will discuss the sequence of events occurring in each stage in the rest of this section.

A sequence of decisions is made before the selling season. At first, the retailer chooses an offer set $S \subseteq N$. Given the offer set S, the retailer determines the selling price p for each selected product and allocates shelf space to each selected product at the local store. After setting prices and allocating shelf space, the retailer places an order and pays a regular wholesale price for each product. Delivery of orders also occurs before the selling season. The variants in set N_1 are sent to the local store: part

of the order is used to fill the shelf space; and the remainder is stored in the backroom. The variants in the set N_2 are sent directly to the warehouse and stored there.

During the selling season, the stock on hand of each product is depleted by the demand during the selling season, which is a realization from a distribution whose density function is determined by SP. At the local store, if demand for product i is greater than the sum of its display inventory and backroom inventory, then the unsatisfied demand is lost; otherwise, there will be leftover inventory of product i at the end of the selling season. At the online store, if the product $i + n$ is in stock at the warehouse, the retailer guarantees to deliver the order within a certain number of days, and this shipping option is named regular shipping. When the inventory of product $i + n$ in the warehouse is depleted, an online consumer is offered two delivery options: regular shipping and *super-saver* shipping. If the consumer chooses regular shipping, the retailer charges this consumer the regular price and uses a drop-shopping strategy to ship the order immediately. On the other hand, if super-saver shipping is chosen, the consumer is charged a discounted price, but the retailer does not ship the order until the end of the selling season.

After the selling season, the retailer fulfills the *super-saver* shipping orders at the lowest cost. The retailer checks the stock level of product i first: if any inventory of product i is left over at the local store, then this leftover inventory of product i is used to fulfill the order of product $i + n$; otherwise, the retailer again uses the drop-shipping strategy to satisfy the outstanding orders. This strategy achieves the lowest cost because we assume that manufacturers charge the retailer higher wholesale prices for drop-shipping items, but there is no extra cost associated with shipping from its own inventory. After delivering super saving shipping items, the retailer disposes of any remaining inventory at a salvage cost.

4.2.5 Demand Models

As mentioned, consumers are categorized within different market segments, and we further assume that each consumer segment is only interested in a subset of the entire product set. Liu and van Ryzin [68] called this set the consumer's *consideration set*. We assume that each consumer belongs to one of K market segments, and each segment k is characterized by one consideration set $C_k \subset N \times \{1, \dots, P\}$. For example, segment-$l$ consumers are low fare walk-in consumers, which means they only shop at the local store and only purchase products with a price less than ρ. As a result, their consideration set C_l can be described as $C_l = \{(i, p) | i \in N_1 , p \leq \rho\}$. As in Bront, Méndez-Díaz and Vulcano [30], we allow for overlapping segments, i.e., we permit $C_k \cap C_l \neq \emptyset$ for $k \neq l$.

In this paper, we use a multinomial logit model (MNL) as the demand model. The MNL model is a utility model that is commonly used in the economics and marketing literature. For a brief description of the MNL, see Anderson et al. [11] or van Ryzin and Mahajan [95]. Under the MNL model, the probability that a segment-k consumer chooses pseudo-product (i, p) from $SP \cap C_k$ is defined by a preference matrix v_k, where the component (i, p) of v_k, v_{kip}, is a segment-k consumer's "preference weight" for pseudo-product (i, p). This matrix, together with the no-purchase preference v_{k0}, determines the probability $P_{kip}(SP)$ as follows:

$$P_{kip}(SP) = \frac{v_{kip}}{\sum_{(j,q) \in C_k \cap SP} v_{kjq} + v_{k0}} \qquad \forall i, p, k \ . \tag{4–1}$$

If $(i, p) \notin SP$ or $(i, p) \notin C_k$, then $v_{kip} = 0$. Moreover, v_{kip} is increasing in consumer utility, so a high value of v_{kip} corresponds to a pseudo-product with a higher expected utility.

One advantage of the MNL model is that it allows one to easily incorporate marketing variables such as prices and promotions into the choice model. In our study, pseudo-products (i, p_1) and (i, p_2) are identical products, and the only difference is that $p_1 \neq p_2$. If $p_1 < p_2$, then it is natural to assume that for some consumer segment

k, we will have $v_{kip_1} > v_{kip_2}$. In addition, we can use simple constraints in our model to reflect the fact that pseudo-products (i, p_1) and (i, p_2) may not be included in a set SP at the same time.

Despite its advantages, there are two major shortcomings of the MNL model. One of these shortcomings stems from its Independence of Irrelevant Alternatives (IIA) property. To illustrate this property, consider two distinct consideration sets S_1 and S_2, and two distinct products i and j. The MNL formula implies that

$$\frac{P_i(S_1)}{P_j(S_1)} = \frac{P_i(S_2)}{P_j(S_2)} = \frac{v_i}{v_j} \quad . \tag{4–2}$$

Thus, the relative preferences of products i and j are independent of the composition of the offer sets S_1 and S_2. The IIA property would not hold in the case where the consideration set could be divided into subsets such that products within one subset are more similar to each other than across subsets. For example, suppose there is a smart phone maniac who has the same probability of purchasing a Samsung Galaxy III or an iPhone 4S, i.e., $P\{\text{Samsung}\} = P\{\text{iPhone}\} = \frac{1}{2}$. There are two models of iPhone 4S that are identical except for their colors, black or white. Assume that the maniac is indifferent about the color of the phone he purchases. If the offer set is $\{$Samsung Galaxy III, black iPhone 4S, white iPhone 4S$\}$, then one would intuitively expect that $P\{\text{Samsung}\} = \frac{1}{2}$ and $P\{\text{black iPhone}\} = P\{\text{white iPhone}\} = \frac{1}{4}$. However, the MNL model implies that that $P\{\text{Samsung}\} = P\{\text{black iPhone}\} = P\{\text{white iPhone}\} = \frac{1}{3}$.

The second shortcoming of the MNL model is that it is unable to fully capture substitution behavior. Kök et al. [61] showed that it is not possible under the MNL model to have two categories with the same preference weight but different substitution rates.

We use a binary vector $x \in \{0, 1\}^{2nP}$ to characterize the set SP, i.e., $x_{ip} = 1$ if $(i, p) \in SP$, and $x_{ip} = 0$ otherwise. We can then express (4–1) in terms of the binary

variables x_{ip} for the MNL choice model:

$$P_{kip}(SP) = P_{kip}(x) = \frac{v_{kip}x_{ip}}{\sum_{(j,q)\in C_k} v_{kjq}x_{jq} + v_{k0}} \qquad \forall i, p, k \ . \tag{4–3}$$

4.2.6 Chance-Constrained Formulation

We first define the notation we will use in formulating a chance-constrained formulation of the product selection and inventory allocation problem. As discussed earlier, we use i and j to index products, p to index prices and k to index consumer segments. We assume that shelf space has three dimensions, but we only use width to measure shelf space. One *facing* of product i consumes α_i units of shelf width, where the total width of the shelf space is A units. The height and depth of shelf determine the capacity of one *facing*, and each facing of product i can hold γ_i units of product i. Letting w_{ip} denote the number of facings allocated to pseudo-product (i, p), then $\alpha_i w_{ip}$ denotes the shelf width consumed by pseudo-product (i, p), and $\gamma_i w_{ip}$ denotes the maximum number of units (i, p) that can be held on the shelf.

For convenience, we measure backroom and warehouse capacity using a single dimension, where B denotes the store's backroom inventory capacity, B^w denotes the warehouse capacity, and β_i denotes the amount of this capacity consumed per unit of product i. If y_{ip} units of pseudo-product (i, p) are ordered, where $i \in N_1$, then the number of units that must be allocated to the backroom equals $y_{ip} - \gamma_i w_{ip}$. As a result, the total amount of backroom capacity consumed by pseudo-product (i, p) equals $\beta_i(y_{ip} - \gamma_i w_{ip})$.

We let I_{ip} denote the number of units of pseudo-product (i, p) inventory remaining at the end of the selling season, while s_{ip} denotes the amount of demand that exceeds the order quantity y_{ip}. If $i \in N_1$, then s_{ip} is the number of lost sales at the store for pseudo-product (i, p); if $i \in N_2$, then s_{ip} corresponds to the total number of units either drop shipped from the manufacturer or transferred from the local store. Let us further assume that if consumers are offered regular shipping and super-saver shipping for

pseudo-product (i, p) at the same time, then on average ϕ_{ip} of them choose regular shipping and $1 - \phi_{ip}$ of them choose super-saving shipping, where $\phi_{ip} \in [0, 1]$.

We also let y_{ip}^d denote the number of units of pseudo-product (i, p) drop shipped from the manufacturer directly to the consumer, while y_{ip}^t denotes the number of units of pseudo-product (i, p) transferred from leftover inventory of product $i - n$. Note that regular shipping of out-of-stock pseudo-product (i, p) requires using drop-shipping, but drop-shipping may be also used at the end of selling season if the remaining inventory is insufficient to satisfy orders for super-saver shipping. This implies y_{ip}^d is at least as large as $\phi_{ip} s_{ip}$. Note that I_{ip}, s_{ip}, y_{ip}^d and y_{ip}^t are random variables that depend on the product order quantities and consumer demands.

Let r_{ip} denote the unit retail price of pseudo-product (i, p). Let c_i and f_i denote, respectively, the unit wholesale price and fixed cost associated with ordering product i, and let g_i and h_i denote, respectively, the unit shortage cost (lost sales or compensation for super-saver shipping) and unit salvage value for product i. Let c_i^d denote the unit drop-shipping wholesale price associated with product i, and note that it is intuitive to have $c_i^d > c_i$. We assume that each time that shelf inventory for product i is replenished from the backroom inventory, a cost of m_i per unit replenished is incurred.

In order to formulate the retailer's decision problem, we need to characterize a function that determines whether any shortages have occurred. Let ξ_k denote a random number of consumers within segment k that arrive during the selling season. Then, because $P_{kip}(x)$ gives the proportion of segment k consumers that demand pseudo-product (i, p) when the set x is offered, $\sum_{k=1}^{K} \xi_k P_{kip}(x)$ gives the total number of demands for pseudo-product (i, p) during the selling season. We define $G(x, y, \xi) = \max_{1 \le i \le n, 1 \le p \le P} \left\{ \sum_{k=1}^{K} \xi_k P_{kip}(x) - y_{ip} \right\}$. Then, if $G(x, y, \xi)$ is less than or equal to zero, this implies that no shortages have occurred, and we can write the probability of no shortages as $P\left(G(x, y, \xi) \le 0\right)$.

In order to compute the expected profit, we first define the function

$$
\begin{aligned}
Q(w,x,y,\xi) = \sum_{i\in N_1}\sum_{p=1}^{P}\Bigg\{ & r_{ip}\left[\sum_{k=1}^{K}\xi_k P_{kip}(x) - s_{ip}\right] + h_i\left(l_{ip} - y^t_{(i+n)p}\right) - g_i s_{ip} \\
& - m_i\left[\sum_{k=1}^{K}\xi_k P_{kip}(x) - s_{ip} - \gamma_i w_{ip}\right]^+ \Bigg\} \\
& + \sum_{i\in N_2}\sum_{p=1}^{P}\left\{ r_{ip}\sum_{k=1}^{K}\xi_k P_{kip}(x) + h_i l_{ip} - g_i(1-\phi_{ip})s_{ip} - c_i^d y^d_{ip}\right\} \quad . \qquad (4\text{–}4)
\end{aligned}
$$

where $l_{ip} = \left[y_{ip} - \sum_{k=1}^{K}\xi_k P_{kip}(x)\right]^+$, $s_{ip} = \left[\sum_{k=1}^{K}\xi_k P_{kip}(x) - y_{ip}\right]^+$ and $y^t_{ip} = \min\{\sum_{p=1}^{P} l_{(i-n)p}, (1-\phi_{ip})s_{ip}\}$ and $y^d_{ip} = \max\{s_{ip} - y^t_{ip}, \phi_{ip}s_{ip}\}$. Next, letting $\mathcal{Q}(w,x,y) = E_\xi[Q(w,x,y,\xi)]$, we formulate the retailer's inventory planning problem as follows:

$$
\theta^* = \max \mathcal{Q}(w,x,y) - \sum_{i\in N}\sum_{p=1}^{P}(f_i x_{ip} + c_i y_{ip}) \qquad (4\text{–}5)
$$

$$
\text{s.t.} \sum_{i\in N_1}\sum_{p=1}^{P}\alpha_i w_{ip} \le A \qquad (4\text{–}6)
$$

$$
\sum_{i\in N_2}\sum_{p=1}^{P}\beta_i y_{ip} \le B^w \qquad (4\text{–}7)
$$

$$
\sum_{i\in N_1}\sum_{p=1}^{P}\beta_i(y_{ip} - \gamma_i w_{ip}) \le B \qquad (4\text{–}8)
$$

$$
\sum_{p=1}^{P} x_{ip} \le 1 \qquad \forall i \in N \qquad (4\text{–}9)
$$

$$
\gamma_i w_{ip} \le y_{ip} \qquad \forall i \in N_1, p = 1,\dots,P \qquad (4\text{–}10)
$$

$$
y_{ip} \le M x_{ip} \qquad \forall i \in N, p = 1,\dots,P \qquad (4\text{–}11)
$$

$$
P\left(G(x,y,\xi) \le 0\right) \ge 1 - \epsilon \qquad (4\text{–}12)
$$

$$
x_{ip} \in \{0,1\}, w_{ip} \ge 0, y_{ip} \ge 0 \qquad \forall i \in N, p = 1,\dots,P \quad . \qquad (4\text{–}13)
$$

The objective function (4–5) is composed of the ordering cost and the expected profit from sales. The first part is the profit at the local store, which is expressed as the difference between the sum of revenues and the sum of salvage cost, stockout cost

and shelf replenishment cost. There are several things to be aware of when computing the profit for product $i \in N_1$: first, a lost sale is possible for a product sold at the local store, which implies the realized sales for pseudo-product (i, p) is $\sum_{k=1}^{K} \xi_k P_{kip}(x) - s_{ip}$; second, part of the leftover inventory of product i may be used to fulfill online consumer demand for product $i + n$, therefore only $I_{ip} - y_{ip}^t$ units will realize the salvage value h_i; and finally, the shelf replenishment cost will only be charged if the realized demand is greater than the on-shelf inventory, and the replenishment cost is linear in the number of units of product replenished. The other part of the profit equation is from the online store, and it is the difference between the sum of revenues and the sum of salvage cost, compensation for super-saver shipping, and drop-shipping cost. Note that there are no lost sales for online products and compensation is only paid to consumers who choose super-saver shipping.

Constraint (4–6) ensures that shelf capacity A at the local store is not violated. Constraint (4–7) states that the capacity allocated to online products is bounded above by B^w. Constraint (4–8) states that the order quantities of products sold at the local store beyond the display inventory are limited by the backroom capacity B at the local store. Constraint (4–10) guarantees that if pseudo-product (i, p) is not ordered at the local store, then no shelf facing capacity is allocated to pseudo-product (i, p); it also ensures that the order quantity is large enough to fill the allocated shelf space. If pseudo-product (i, p) is not included in the local/online store assortment, then no pseudo-product (i, p) inventory is ordered at the local/online store, and constraint (4–11) captures this assumption. Finally, the nonnegativity and integrality constraints are provided in (4–13). Constraint (4–12) indicates that $G(w, x, y, \xi)$ should be non-positive with a probability of at least $1 - \epsilon$, which implies that there is at most an ϵ probability of a shortage for any product at the local store.

4.3 Solution Procedure

The first difficulty in solving our problem is that the MNL model includes nonlinear terms which need to be linearized in order to convert our problem to a form that is well behaved in terms of chance-constrained stochastic programs. Standard linearization techniques are used to convert the expressions under the MNL demand models to linear expressions.

4.3.1 MNL Model Linearization

In this section, we follow the procedure used by Méndez-Díaz and Vulcano [30] to linearize our MNL model. By defining the variables

$$\pi_k = \frac{1}{\sum_{(j,q)\in C_k} v_{kjq} x_{jq} + v_{k0}}, \qquad \forall k = 1 \dots, K,$$

the probability $P_{kip}(x)$ in the MNL model can be rewritten as

$$P_{kip}(x) = v_{kip}\pi_k x_{ip} \qquad \forall i \in N, p = 1, \dots, P, k = 1, \dots, K$$

$$\sum_{(j,q)\in C_k} v_{kjq}\pi_k x_{jq} + v_{k0}\pi_k = 1 \qquad \forall k = 1, \dots, K \ .$$

Wu [99] proved that the polynomial mixed $0-1$ term $u_{kip} = \pi_k x_{ip}$, where x_{ip} is a binary variable and π_k is a nonnegative continuous variable, can be represented by the following linear inequalities: (i) $\pi_k - u_{kip} \leq M_1 - M_1 x_{ip}$; (ii) $u_{kip} \leq \pi_k$; (iii) $u_{kip} \leq M_1 x_{ip}$; and (iv) $u_{kip} \geq 0$, where M_1 is a large number greater than $\frac{1}{\min_{k,i,p}\{v_{kip}\}}$. By applying this result,

we obtain the following reformulation of the MNL model:

$$P_{kip}(x) = v_{kip}u_{kip} \qquad \forall i \in N, p = 1, \dots, P, k = 1, \dots, K \tag{4–14}$$

$$\sum_{(j,q)\in C_k} v_{kjq}u_{kjq} + v_{k0}\pi_k = 1 \qquad \forall k = 1, \dots, K \tag{4–15}$$

$$\pi_k - u_{kip} \leq M_1 - M_1 x_{ip} \qquad \forall (i,p) \in C_k, k = 1, \dots, K \tag{4–16}$$

$$u_{kip} \leq \pi_k \qquad \forall (i,p) \in C_k, k = 1, \dots, K \tag{4–17}$$

$$u_{kip} \leq M_1 x_{ip} \qquad \forall (i,p) \in C_k, k = 1, \dots, K \tag{4–18}$$

$$u_{kip} \geq 0 \qquad \forall (i,p) \in C_k, k = 1, \dots, K \ . \tag{4–19}$$

4.3.2 Sample Average Approximation

The second difficulty in our problem is that our model contains both chance-constrained and two-stage stochastic program features. In general, evaluating the objective function and checking solution feasibility of this type of problem is not easy, and the feasible region defined by the chance constraint generally is not convex. Wang and Guan [97] call this type of program a CCTS program, and they developed a combined sample average approximation (SAA) algorithm to solve the CCTS program effectively.

A sample average approximation of our true problem is obtained by replacing the true distribution of demands by an empirical distribution corresponding to a random sample. In this study, we assume that consumers in any segment k arrive according to a Poisson distribution with rate λ_k. In the previous section, we defined ξ_k as a random number of consumers in segment k that arrive during the selling season, and now we further define $\xi \equiv \{\xi_1, \dots, \xi_K\}$ as the random demand vector. Now let ξ^1,...,ξ^L be an independent identically distributed (i.i.d) sample of L realizations of ξ generated by

Monte Carlo simulation, and consider the following problem associated with this sample:

$$\max\ \hat{\theta}_L(u,w,x,y) = -\sum_{i=1}^{2n}\sum_{p=1}^{P}(f_i x_{ip} + c_i y_{ip}) + \frac{1}{L}\sum_{l=1}^{L} Q(u,w,x,y,\xi^l)$$

$$\text{s.t. } (4\text{–}6)\text{-}(4\text{–}11),\ (4\text{–}13),\ (4\text{–}14)\text{-}(4\text{–}19)$$

$$\frac{1}{L}\sum_{l=1}^{L}\mathbb{1}_{(0,\infty)}(G(u,w,x,y,\xi^l)) \le \epsilon\ . \tag{4–20}$$

The above optimization problem is usually referred as a sample average approximate (SAA) problem of the true problem. In contrast to the true problem, a sample average function $\frac{1}{L}\sum_{l=1}^{L} Q(u,w,x,y,\xi^l)$ is used in place of the expected value function $E_\xi[Q(u,w,x,y,\xi)]$, and a function $\frac{1}{L}\sum_{l=1}^{L}\mathbb{1}_{(0,\infty)}(G(u,w,x,y,\xi^l)) \le \epsilon$ is used to estimate the probability of no shortage[2] . Notice that the SAA problem is also a CCTS program, as is the original problem, but with a different demand distribution.

Define $\Lambda \equiv (u,w,x,y)$ as the vector of decision variables. Assume that Λ^* is an optimal solution for the true problem and $\hat{\Lambda}_L$ is an optimal solution for the SAA problem. Let θ^* represent the optimal objective value of the true problem and $\hat{\theta}_L$ represent the optimal objective value of the SAA problem. Wang and Guan [97] showed that $\hat{\theta}_L \to \theta^*$ and $D(\hat{\Lambda}_L, \Lambda^*) \to 0$ with probability one as $L \to \infty$, where $D(\hat{\Lambda}_L, \Lambda^*)$ represents the distance between $\hat{\Lambda}_L$ and Λ^*.

4.3.2.1 Solution validation

A candidate solution $\hat{\Lambda}_L$ of the SAA problem is not necessary an optimal solution for the true problem. Actually, $\hat{\Lambda}_L$ may not even be feasible for the true problem. As a result,

[2] In this study, we define the indicator function $\mathbb{1}_{(0,\infty)}(t) : \mathbb{R} \to \mathbb{R}$ as follows

$$\mathbb{1}_{(0,\infty)}(t) = \begin{cases} 1 & \text{if } t > 0 \\ 0 & \text{if } t \le 0\ . \end{cases}$$

That is, $\frac{1}{L}\sum_{l=1}^{L}\mathbb{1}_{(0,\infty)}(G(u,w,x,y,\xi^l))$ is equal to the proportion of realizations in which a stockout occurs.

after obtaining a candidate solution $\hat{\Lambda}_L$ of the SAA problem, it is important to validate its quality as a solution of the true problem.

The solution validation starts with the verification of feasibility of $\hat{\Lambda}_L$. Given $\hat{\Lambda}_L$, let

$$q(\hat{\Lambda}_L) = P\left(G(\hat{\Lambda}_L, \xi) \leq 0\right) \quad .$$

Because it is difficult to compute $q(\hat{\Lambda}_L)$ exactly, in order to check the feasibility of $\hat{\Lambda}_L$, we followed the method developed by Ahmed and Shapiro [3, 4] to construct an approximate $(1-\beta)$ confidence upper bound on $q(\hat{\Lambda}_L)$. First, generate an i.i.d sample $\xi^1, \ldots, \xi^{\bar{L}}$ which should be independent of the sample producing the candidate solution $\hat{\Lambda}_L$. Given this newly generated sample and $\hat{\Lambda}_L$, let

$$\hat{q}_{\bar{L}}(\hat{\Lambda}_L) = \frac{1}{\bar{L}} \sum_{l=1}^{\bar{L}} \mathbb{1}_{(0,\infty)}(G(\hat{\Lambda}_L, \xi^l)) \quad .$$

Notice here the size $\bar{L}$ of this sample could be very large, since we do not need to solve any optimization problem using this sample. Moreover, for not "too small" $q(\hat{\Lambda}_L)$ and large $\bar{L}$[3] , $\hat{q}_{\bar{L}}(\hat{\Lambda}_L)$ is not only an unbiased estimator of $q(\hat{\Lambda}_L)$, but the distribution of $q(\hat{\Lambda}_L)$ can also be approximated reasonably well by a normal distribution with mean $\hat{q}_{\bar{L}}(\hat{\Lambda}_L)$ and variance $\hat{q}_{\bar{L}}(\hat{\Lambda}_L)(1-\hat{q}_{\bar{L}}(\hat{\Lambda}_L))/\bar{L}$. As a result, we can construct the following $(1-\beta)$-confidence upper bound on $q(\hat{\Lambda}_L)$:

$$U_{\beta,\bar{L}}(\hat{\Lambda}_L) = \hat{q}_{\bar{L}}(\hat{\Lambda}_L) + \Phi^{-1}(1-\beta)\sqrt{\hat{q}_{\bar{L}}(\hat{\Lambda}_L)(1-\hat{q}_{\bar{L}}(\hat{\Lambda}_L))/\bar{L}} \ ,$$

where Φ^{-1} is the inverse of the cumulative distribution of standard normal distribution. It is easy to see that if $U_{\beta,\bar{L}}(\hat{\Lambda}_L)$ is less than ϵ, then we are $(1-\beta)$ confident that $\hat{\Lambda}_L$ is feasible for the true problem and $\hat{\theta}_{\bar{L}}(\hat{\Lambda}_L)$ is a lower bound for θ^*.

[3] To be on the safe side, one should require $\bar{L} \times q(\hat{\Lambda}_L)$ to be greater than or equal to 5.

After checking the feasibility of solution $\hat{\Lambda}_L$, the next step of the solution validation is to estimate the optimality gap $\theta^* - \hat{\theta}_{\bar{L}}(\hat{\Lambda}_L)$, and we again followed the method provided by Ahmed and Shapiro [3, 4] to obtain a upper bound for the objective value θ^*. In this method, we generated M independent samples $\xi^{1,m}, \dots, \xi^{L,m}$, $m = 1, \dots, M$, each of size L. For these M samples, we picked the Tth largest optimal value as the approximate upper bound for θ^* with confidence level $(1-\beta)$, where T is calculated as described in [4]. Then the optimality gap can be estimated by the difference between this approximated upper bound and $\hat{\theta}_{\bar{L}}(\hat{\Lambda}_L)$.

4.3.2.2 SAA formulation

To solve the SAA problem using a commercial solver, the chance constraint (4–20) should be reformulated as follows:

$$\sum_{k=1}^{K} \xi_k^l P_{kip}(x) - y_{ip} \leq M_2 z_l \qquad \forall i \in N_1, p = 1, \dots, L, l = 1, \dots, L$$

$$\sum_{l=1}^{L} z_l \leq L \times \epsilon$$

$$z_l \in \{0, 1\} \qquad \forall l = 1, \dots, L \ ,$$

where M_2 is a large number. After reformulation, the SAA problem becomes a mixed integer program with a large number of binary variables. In the SAA literature, in order to speed up the algorithm, star-inequalities are often used to strength the formulation of the SAA problem. However, for our problem, star-inequalities are not applicable.

When there is only one consumer segment under consideration (i.e., $K = 1$), after taking a sample of L realizations, the demand realizations can be sorted in nonincreasing order, and we assume that ξ^l are indexed such that $\xi^1 \geq \dots \geq \xi^L$. The constraint $\sum_{l=1}^{L} z_l \leq L \times \epsilon$ implies that we cannot have $z_l = 1$ for all $l = \lfloor \epsilon \times L \rfloor + 1, \dots, L$. We can therefore tighten the SAA formulation by fixing

$$z_l = 0, \forall l = \lfloor \epsilon \times L \rfloor + 1, \dots, L \ .$$

It can be observed that this sorting method does not work for the general case in which there is more than one consumer segment, because each demand realization ξ^l is a vector which is composed of more than one component. At this stage, we could not find strong inequalities for the general case of the SAA formulation. As a result, in the next section, we propose a heuristic for solving our problem in general.

4.4 Greedy and Local Search Algorithm

In the greedy and local search algorithm, the set SP is constructed from scratch (an empty set), choosing at each iteration the pseudo-product bringing the "highest" additional profit.

Given a set SP, the original CCTS problem with both linear and binary variables can be reduced to another CCTS problem with only linear variables, which can be written as follows

$$\theta^*(SP) = \max - \sum_{(i,p)\in SP} (f_i + c_i y_{ip}) + E_\xi \left[Q\left(SP, w, y\right)\right] \tag{4–21}$$

$$\text{s.t.} \sum_{(i,p)\in SP_1} \alpha_i w_{ip} \leq A \tag{4–22}$$

$$\sum_{(i,p)\in SP_2} \beta_i y_{ip} \leq B^w \tag{4–23}$$

$$\sum_{(i,p)\in SP_1} \beta_i (y_{ip} - \gamma_i w_{ip}) \leq B \tag{4–24}$$

$$\gamma_i w_{ip} \leq y_{ip} \qquad \forall (i,p) \in SP_1 \tag{4–25}$$

$$P\left(G(SP, y, \xi) \leq 0\right) \geq 1 - \epsilon \tag{4–26}$$

$$w_{ip} \geq 0, y_{ip} \geq 0 \qquad \forall (i,p) \in SP \ , \tag{4–27}$$

where

$$Q\left(SP, w, y\right) = \sum_{(i,p)\in SP_1} \left\{ r_{ip}\left[\sum_{k=1}^{K} \xi_k P_{kip}(SP) - s_{ip}\right] + h_i\left(I_{ip} - y^t_{(i+n)p}\right) - g_i s_{ip} \right.$$
$$\left. - m_i\left[\sum_{k=1}^{K} \xi_k P_{kip}(SP) - s_{ip} - \gamma_i w_{ip}\right]^+ \right\}$$
$$+ \sum_{(i,p)\in SP_2} \left\{ r_{ip}\sum_{k=1}^{K} \xi_k P_{kip}(SP) + h_i I_{ip} - g_i(1-\phi_{ip})s_{ip} - c_i^d y_{ip}^d \right\} \quad ,$$

and $G(SP, y, \xi) = \max_{(i,p)\in SP_1}\left\{\sum_{k=1}^{K} \xi_k P_{kip}(SP) - y_{ip}\right\}$, and $SP_j \equiv \{(i,p)|(i,p) \in SP, i \in N_j\}, \forall j = 1, 2$. The above problem can be considered as a restricted version of the original problem. It is easier to solve this restricted problem than the original one, because the restricted problem does not include any binary variables, and the probability terms $P_{kip}(SP)$ are constant. We still use the SAA algorithm to solve this restricted problem but with two modifications. First, the star-inequalities are added to the SAA problem to speed up the computation. To do this, we introduce a new set of binary variables $\{z_{ipl} \in \{0,1\} : i \in N; p = 1, \dots, P; l = 1, \dots, \lfloor L \times \epsilon \rfloor\}$ and define $d_{ipl} = \sum_{k=1}^{K} \xi_k^l P_{kip}(SP)$. We sort the d_{ipl} in decreasing order of magnitude, i.e., $d_{ip1} \geq d_{ip2} \dots \geq d_{ip\lfloor L\times\epsilon\rfloor}$. Then the star-inequalities can be written as follows:

$$\begin{aligned}
& z_{ipl} - z_{ip(l+1)} \geq 0 && \forall (i,p) \in SP, l = 1, \dots, \lfloor L \times \epsilon \rfloor \\
& z_l - z_{ipl} \geq 0 && \forall (i,p) \in SP, l = 1, \dots, \lfloor L \times \epsilon \rfloor \\
& y_{ip} + \sum_{l=1}^{\lfloor L\times\epsilon\rfloor} (d_{ipl} - d_{ip(l+1)}) z_{ipl} \geq z_{ip1} && \forall (i,p) \in SP \\
& \sum_{l=1}^{L} z_l \leq L \times \epsilon \\
& z_l, z_{ipl} \in \{0,1\} && .
\end{aligned}$$

Luedtke et al. [69] proved that the star-inequalities can strengthen the formulation of the SAA problem. Second, the goal of this heuristic is to generate a good feasible

solution and no upper bound needs to be generated, so the solution validation step that generates an upper bound is skipped after getting a $(1-\beta)$-confidence lower bound $\hat{\theta}(SP)$ of $\theta^*(SP)$. The greedy and local search algorithm can be summarized as follows.

1. Set $SP = \emptyset$.
2. Set $(i, p) = \text{argmax}\{\hat{\theta}(SP \cup (j, q)), (j, q) \notin SP\}$
3. If $\hat{\theta}(SP \cup (i, p)) \leq \hat{\theta}(SP)$, then stop.
4. Otherwise, set $SP = SP \cup (i, p)$, and return to 2.

4.5 Numerical Experiments

This section presents computational results for the combined SAA algorithm and the heuristic approach for solving our CCTS problem. First, we demonstrate the convergence property of the combined SAA algorithm for our problem. Second, we compare the performance of these two different solution approaches. In addition to these, we also analyzed the impacts of different parameters and consumer behaviors on the assortment and pricing decisions. We implemented both approaches in the C# programming language, with the mixed integer SAA problem solved using ILOG®'s CPLEX® 12.5 solver with Concert Technology. We performed all tests on a computer with an Intel® Dual Core 1.7 GHz and 6 GB memory.

4.5.1 Convergence Tests

To demonstrate the convergence property of the combined SAA algorithm, we used one problem set with two consumer segments (i.e., $K = 2$), ten products (i.e.,$|N_1| = |N_2| = 10$) and three price levels (i.e., $P = 3$). We tested five randomly generated instances for this problem set. For each problem instance, we set different values for L, $\bar{L}$ and M to show that optimality gap decreases as the sample size increases. We used $(1-\beta) = 99\%$ as our estimation confidence level.

Table 4-1 summarizes the distributions used in generating parameters in our computational study. For each problem instance, the product widths α_i, volumes β_i, unit wholesale costs c_i, consumer's average arrival rates γ_k and consumer's preferences

Table 4-1. Parameter distributions used in computational tests.

Product i's width α_i $U(1,3)$	Product i's volume β_i $U(1,9)$
Retailer's unit wholesale cost, c_i $U(0.5,5)$	Consumer's average arrival rates λ_k $U(1000,5000)$
Retailer's fixed cost, f_i $c_i \times U(10,100)$	Consumer's preferences at regular price v_{ki1} $U(0,15)$
Retailer's unit price r_i $c_i \times U(1.3,1.7)$	Consumer's preferences at price p v_{kip} $v_{ki1} \times 1.3^{p-1}$
Retailer's unit inventory salvage value h_i $c_i \times U(0.5,0.9)$	Retailer's unit shortage cost, g_i $c_i \times U(0.1,0.5)$
Retailer's unit drop-shipping value c_i^d $c_i \times U(1.1,1.5)$	Retailer's unit replenished cost, m_i $c_i \times U(0.01,0.1)$
Shelf space length A $\bar{\alpha} \times \lambda \times U(0.01,0.5)$	Backroom inventory capacity B $\bar{\beta} \times \lambda \times U(0.2,0.5)$
Warehouse inventory capacity B^w $\bar{\beta} \times \lambda \times U(0.2,0.5)$	Product i's facing capacity γ_i $\lfloor \frac{9 \times \alpha_i}{\beta_i} \rfloor$
Local store service level $1-\epsilon$ $1-U(0.01,0.2)$	

at regular price v_{ki1} were first generated from some uniform distributions. Second, we calculated the average width $\bar{\alpha}$ and the average volume $\bar{\beta}$ over all products, and total expected demand λ over all consumer segments. Finally, based on these generated values and calculated values, the values of other parameters were generated based on additional independent continuous uniform distributions. As in Chapter 3, we let $U(l,u)$ denote the continuous distribution with lower bound l and upper bound u.

Table 4-2 summarizes the results for instance 1, and tables for other instances are presented in Appendix E. All tables show that a larger value of L results in a smaller relative optimality gap, which shows empirically that the optimal solution

Table 4-2. Computational results I - instance 1.

$(M, \bar{L})$	L	LB	UB	Gap = $\frac{LB}{UB} \times 100$ (%)	Time(s)
(25,1000)	20	32810	33105	0.90	114
	50	32813	32867	0.16	224
	100	32807	32852	0.13	448
(25,2000)	20	32810	32877	0.51	220
	50	32814	32860	0.14	307
	100	32821	32858	0.11	564
(50,1000)	20	32812	32948	0.41	229
	50	32809	32861	0.15	476
	100	32830	32850	0.06	828
(50,2000)	20	32820	32934	0.34	434
	50	32821	32866	0.13	594
	100	32831	36848	0.05	1083

indeed converges as L increases. However, the best feasible solution (i.e., LB) may not necessarily increase as L increases. This is because, as L increases, the feasible region of the approximation becomes smaller which would lead to more conservative solutions. This observation is consistent with the findings of Luedtke and Ahmed [69]. Moreover, for a fixed value L, the relative optimality gap decreases as the value of $M \times \bar{L}$ increases.

4.5.2 Comparison of Solution Methods

To benchmark the performance of the SAA algorithm with the greedy and local search heuristic, we tested our solution methods using twelve problem sets. Each problem set corresponds to a fixed number of consumer segments, number of products and number of price levels, $(K, |N_1|, |N_2|, P)$, where $K \in \{1, 3\}$, $|N_1|, |N_2| \in \{5, 10\}$ and $P \in \{2, 3, 4\}$. We tested twenty randomly generated instances for each combination of $(K, |N_1|, |N_2|, P)$ values, for a total of 240 problem instances. For the SAA algorithm, we set the scenario parameters $L = 100$, $M = 25$, $\bar{L} = 1000$ and $\beta = 1\%$. For the greedy and local search heuristic, we set $L = 1000$.

To compare the SAA algorithm and the greedy and local search heuristic, we consider the running time and relative optimality performance. The results of our tests, averaged over the twenty random problem instances for each combination of

$(K, |N_1|, |N_2|, P)$ values, are presented in Table 4-3. In the table, "SAA" represents the SAA algorithm and "Greedy" represents the greedy and local search heuristic. The table shows that for both methods, the running time increases in L, $\bar{L}$ and M. For the majority of problems, the heuristic runs faster than the SAA algorithm, except for problems with $(|N_1|, |N_2|, P) = (10, 10, 2)$. This is because the running time of the SAA algorithm increases exponentially as the size of of the problem increases, but the running time of the heuristic increases polynomially as the size of the problem increases [4] . As a result, for small problem instances, it is possible that the SAA algorithm may run faster than the heuristic, but we expect that the speed dominance of the heuristic will become more and more obvious as the number of products or the number of price levels increases.

For all problem sets, the relative optimality performance of both methods decreases as the problem size increases. As a myopic method, the heuristic still has an average relative optimality gap under 2%. In conclusion, for large size problem instances, the heuristic has much shorter running time and acceptable optimality performance. As a result, the greedy and local search heuristic can be used as an alternative solution approach for large size problems when computing power is limited or solution time is critical.

4.5.3 Parameter Analysis

The goal of this section is to study how different parameters influence the clicks-and-mortar retailer decisions. For comparison purposes, we used a base parameter setting in this section, as summarized in Table 4-4. For each random problem instance used, α_i, β_i, c_i, λ_k and v_{kp1} are still generated from continuous uniform

[4] Both methods solve a series of mixed integer programs. For the SAA algorithm, the number of mixed integer programs solved is M and the number of binary variables in each mixed integer program is equal to $(|N_1| + |N_2|)P + L$. At the same time, the number of mixed integer programs solved in the heuristic is at most $\frac{(|N_1|+|N_2|+1)(|N_1|+|N_2|)}{2}P$ and the number of binary variables in each program is equal to L.

Table 4-3. Computational results II.

Problem set			SAA		Greedy	
K	$\lvert N_1\rvert = \lvert N_2\rvert$	P	Relative gap (%)	Time	Relative gap (%)	Time
1	5	2	0.12	74	0.66	29
		3	0.17	95	0.91	42
		4	0.20	162	1.11	54
	10	2	0.15	105	0.85	138
		3	0.20	225	0.94	153
		4	0.38	348	1.10	193
3	5	2	0.09	204	0.69	73
		3	0.12	283	0.91	94
		4	0.37	400	1.30	115
	10	2	0.31	261	1.21	284
		3	0.69	472	1.32	317
		4	0.88	878	1.70	364

distributions. However, in contrast to Table 4-1 used in Subsections 4.5.1 and 4.5.2, the values of other parameters are determined by deterministic functions of α_i, β_i, c_i, λ_k or v_{kp1}. Using this base parameter setting, we examined how changes in specific parameter values affect the numerical results. For a parameter of interest, e.g., the retailer's salvage value h_i, we generated three random problem instances for the case of $(K, |N_1|, |N_2|, P) = (2, 10, 10, 3)$ using the base parameter setting, then we fixed the values of all other parameters for each problem instance, and considered how changing the value of the parameter of interest affects the numerical results. That is, for the salvage value h_i, we consider five cases:

1. $h_i = 0.1 \times c_i, i = 1, \dots, |N|$;
2. $h_i = 0.3 \times c_i, i = 1, \dots, |N|$;
3. $h_i = 0.5 \times c_i, i = 1, \dots, |N|$;
4. $h_i = 0.7 \times c_i, i = 1, \dots, |N|$; and
5. $h_i = 0.9 \times c_i, i = 1, \dots, |N|$.

We then computed the following criteria for each case:

- maximum profit θ;

- shelf space utilization $(SU) \equiv \frac{\sum_{i \in N_1} w_i}{A} \times 100\%$;
- backroom capacity utilization $(BU) \equiv \frac{\sum_{i \in N_1} (y_i - \gamma_i w_i)}{B} \times 100\%$;
- warehouse capacity utilization $(WU) \equiv \frac{\sum_{i \in N_2} y_i}{B^w} \times 100\%$; and
- adjusted service level $(AS) \equiv \frac{\sum_{i \in N} [\sum_{k=1}^{K} \xi_k^l P_{kip}(x) - s_{ip}^l]}{\sum_{k=1}^{K} \sum_{l=1}^{L} \xi_k^l}$.

The criterion "adjusted service level" is different from the service level for two reasons: first, it accounts for both the local store and online store, because the online out-of-stock item still induces extra cost, even though the online store always satisfies demands; second, the service level only considers consumers who decide to make a purchase, whereas the "adjusted service level" considers all customer arrivals, even if the consumer purchases nothing.

The figures that follow illustrate the results of these experiments, which lead to the following observations.

In Figure 4-2, we varied the value of the retailer's salvage value h_i and fixed the values of all other parameters. The figure shows that a higher value of h_i implies larger values of the retailer's optimal profit θ, the retailer's warehouse capacity utilization WU and the adjusted service level AS. This is because a higher value of h_i implies a lower cost of remaining inventory at the end of the selling season. As a result, a higher value of h_i usually leads to a larger online order quantity y_i, which utilizes more warehouse capacity and covers more consumer demands. But no relationship can be observed between h_i and the retailer's local inventory space utilization (i.e., BU and SU). This is because the local inventory level is restricted by the service level $1 - \epsilon$, and the 0.95 service level in the base parameter setting leaves little room for increasing local order quantity when value of h_i is high.

In Figure 4-3, we show how performance criteria change with the retailer's unit drop-shipping cost c_i^d. For all problem instances, a higher value of c_i^d implies a lower value of θ. This effect is quite intuitive, because the increased cost of on-line channel

Table 4-4. Parameter distributions used in parameter analysis.

Product i's width α_i $U(1,3)$	Product i's volume β_i $U(1,9)$
Retailer's unit wholesale cost, c_i $U(0.5,5)$	Consumer's average arrival rates λ_k $U(1000,5000)$
Retailer's fixed cost, f_i $c_i \times 50$	Consumer's preferences at regular price v_{ki1} $U(0,15)$
Retailer's unit price r_i $c_i \times 1.5$	Consumer's preferences at price p v_{kip} $v_{ki1} \times 1.3^{p-1}$
Retailer's unit inventory salvage value h_i $c_i \times 0.7$	Retailer's unit shortage cost, g_i $c_i \times 0.3$
Retailer's unit drop-shipping value c_i^d $c_i \times 1.3$	Retailer's unit replenished cost, m_i $c_i \times 0.05$
Shelf space length A $\bar{\alpha} \times \lambda \times 0.2$	Backroom inventory capacity B $\bar{\beta} \times \lambda \times 0.5$
Warehouse inventory capacity B^w $\bar{\beta} \times \lambda \times 0.5$	Product i's facing capacity γ_i $\lfloor \frac{9 \times \alpha_i}{\beta_i} \rfloor$
Local store service level $1 - epsilon$ 0.95	

stockouts reduces the retailer's profit. In addition, for all problem instances, WU and AS increase as the c_i^d increases. This is because a higher value of c_i^d makes it less profitable to drop-ship unsatisfied online consumer orders. As a result, the retailer will increase the inventory level at the warehouse to reduce the probability of stockout at the online store..

Figure 4-4 shows that when the value of ϵ increases, θ also increases. Because a higher value of ϵ implies a larger feasible region for the the chance-constrained two stage program, we expect the retailer will gain a higher profit as a result. We also observed that BU and SU decrease as ϵ increases. This observation is consistent with

the fact that an increase in service level requires an increase in the inventory level, which results in high utilization of local store space.

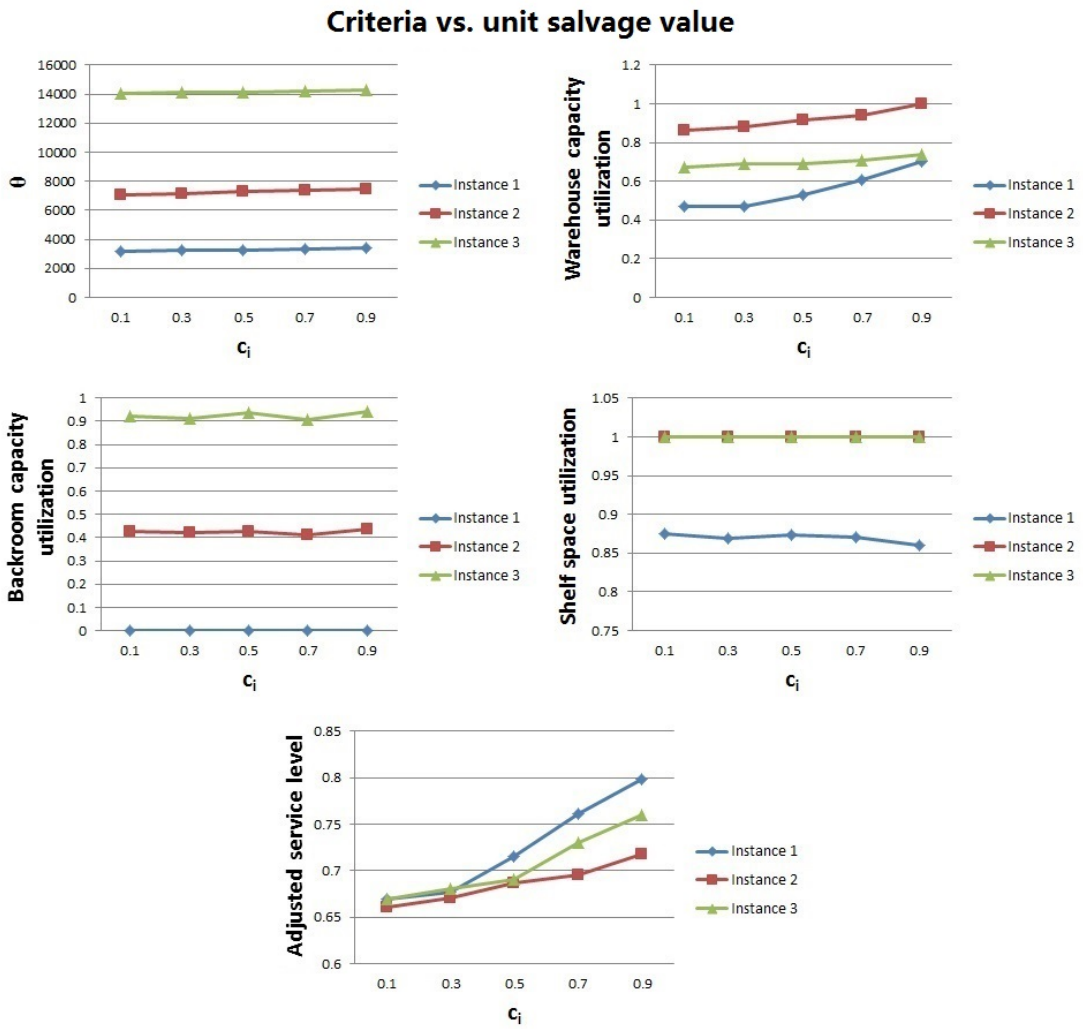

Figure 4-2. Different criteria with h_i at different levels.

4.5.4 Consumer Behavior Analysis

In addition to the above analysis, we are also interested in how consumer purchasing behaviors affect the retailer assortment decisions. To this end, we still used the basic setting summarized in Table 4-4. We generated three random problem instances for the case of $(K, |N_1|, |N_2|, P) = (5, 5, 5, 3)$ using the base parameter setting.

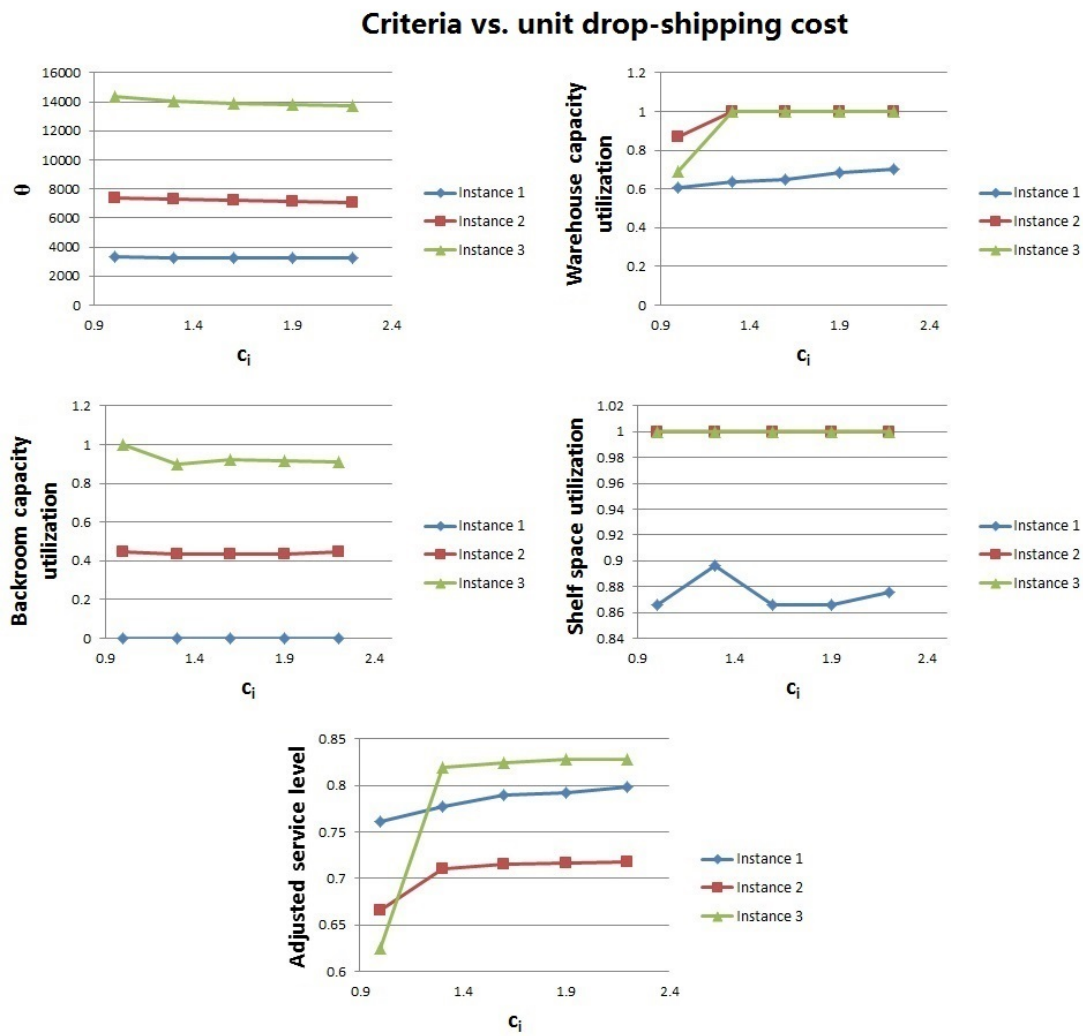

Figure 4-3. Different criteria with c_i^d at different levels.

There is one difference from Subsection 4.5.3: here we assume that all products are identical, i.e., each product has same costs, price and size.

We first studied the effects of the consumer's consideration set, and we considered the following three cases:

1. completely overlapping case: $C_k = \{(i, p) | i \in N, p = 1, \dots, P\}$;
2. partially overlapping case: $C_k = \{(i, p) | i \in N - \{k, k+5\}, p = 1, \dots, P\}$; and
3. completely differentiated case: $C_k = \{(i, p) | i \in \{k, k+5\}, p = 1, \dots, P\}$.

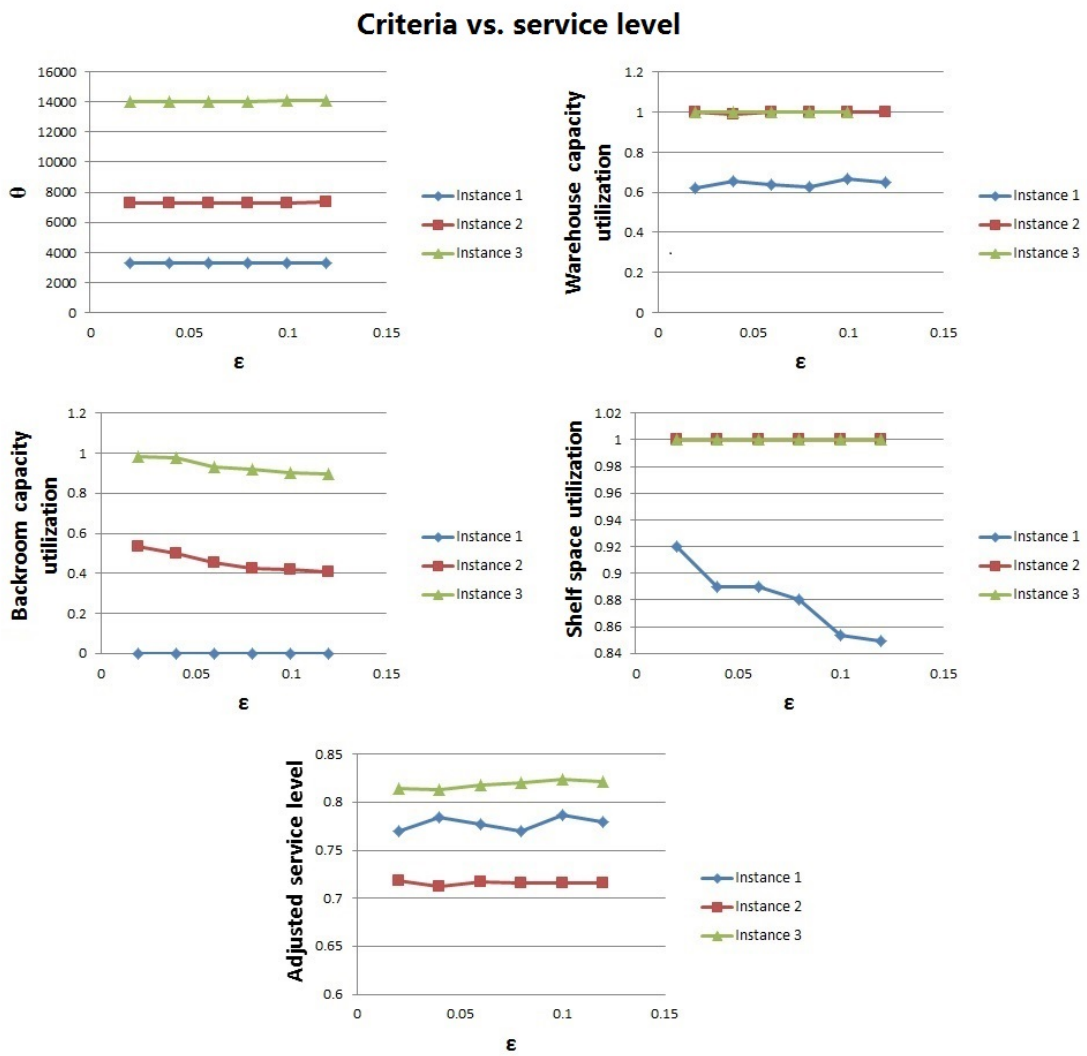

Figure 4-4. Different criteria with ϵ at different levels.

In the completely overlapping case, the consideration sets of any two consumer segments are identical. In contrast, in the completely differentiated case, the consideration sets of any two different consumer segments are mutually exclusive. The partially overlapping case is in between these two cases. The effects can be seen in Figure 4-5 in which Li (Oi) corresponds to the product i at local (online) channel. Before interpreting the results, we first define the retailer *stocking rate* as the ratio of the number of products included in the offer set S to the total number of products in the entire set N. According to this definition, a higher retailer stocking rate implies a wider

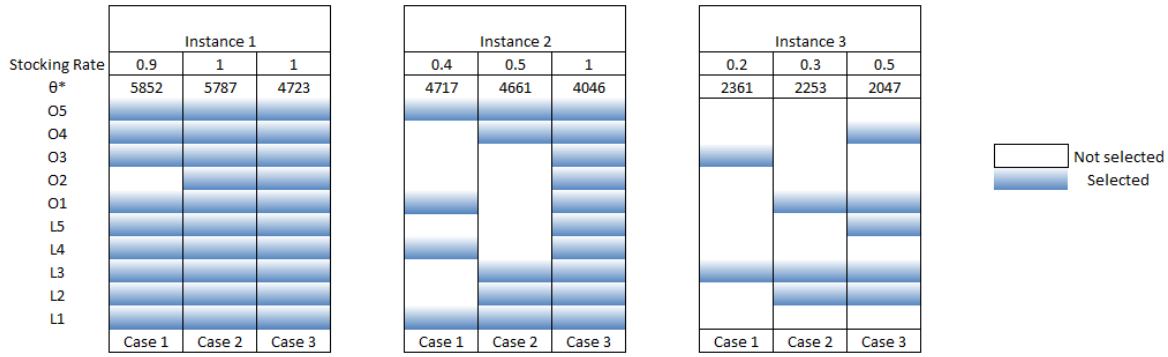

Figure 4-5. Effects of consumer's consideration set on assortment.

product assortment. It should be noted that for all problem instances, the completely overlapping case has the lowest retailer stocking rate and the completely differentiated case has the highest retailer stocking rate. This is because, under the MNL model, increasing the retailer stocking rate increases consumers opportunities to find products that best satisfy their needs and decreases the probability of "no purchase". At the same time, a wider assortment may also increase the within-retailer cannibalization. In the completely overlapping case, adding a product to a nonempty offer set will increase sales cannibalization for sure. On the other hand, in the completely differentiated case, adding a product $i \in N_1$ will not induce cannibalization unless product $i + 5$ is already in the offer set. In summary, a market with higher consumer diversity tends to require a wider product assortment.

In addition to the consumer's consideration sets, we also consider the effects of consumer preference weights on the retailer assortment decisions. To this end, we generated another three random problem instances for the case of $(K, |N_1|, |N_2|, P) = (5, 5, 5, 3)$ using the base parameter setting, and we still assumed that all the products are identical. Moreover, we assumed all the consumer segments are identical and have the entire product set as a consideration set. These two assumptions imply that, for any value of p, the consumer's preference weights v_{kip} are the same for all k and i. For each problem instance, we consider four cases:

1. $v_{ki1} = v_{k0}$;

2. $v_{ki1} = 3v_{k0}$;

3. $v_{ki1} = 5v_{k0}$; and

4. $v_{ki1} = 7v_{k0}$.

The results are presented in Figure 4-6. We noticed that, as the consumer preference weight v_{kip} increases, the retailer stocking rate declines, and the retailer's optimal profit increases. This is because that a low consumer preference weight implies a high "no purchase" utility. van Ryzin and Mahajan [95] pointed out that a high "no purchase" utility represents the existence of many attractive external options, and it is in the retailer's interest to have a wider assortment. On the other hand, a low "no purchase" utility means less external threat, and more threats come from within-retailer cannibalization, so the retailer should limit the breadth of the assortment. Last but not least, a higher value of "no purchase" utility implies a more competitive retail environment, and it is intuitive that the retailer will have lower optimal profit under a more crowded market.

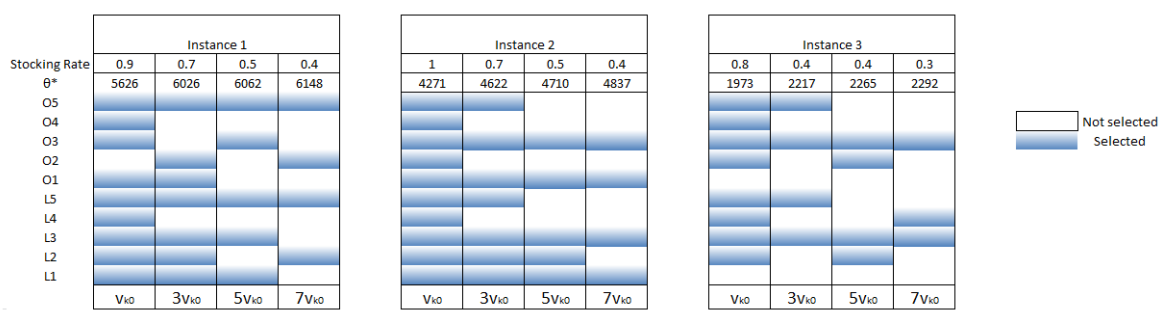

Figure 4-6. Effects of consumer's preference weights on assortment.

CHAPTER 5
CONCLUSIONS

In this chapter, we summarize the research we discussed in Chapter 2 through Chapter 4. We review the employed methodologies and corresponding numerical results, and we also suggest future research directions.

Chapter 2 examines the degree to which the bullwhip effect results from price fluctuations in a two-echelon supply chain with deterministic and price-sensitive demand. We provide numerical evidence that increased system profit can coexist with the bullwhip effect as a result of price promotions if: (i) the supplier judiciously sets the price discount; (ii) there is a sufficient number of impulsive customers who buy the product at the discounted price; (iii) the price discount does not induce a high degree of end-customer forward buying. However, even when the total system profit increases, the retailer takes a disproportionately larger share of this profit gain, while the supplier incurs greater operations costs and tends to observe a marginal profit gain. The numerical experiments also illustrate two mechanisms that cause the bullwhip effect. In addition, this chapter builds a basic structure for future research. For example, we may consider a generalized model with stochastic demand, treating the supplier's price discount d as a decision variable instead of parameterizing on it. We are also interested in issues related to handling multiple products and designing supplier-retailer contracts for discount policies, both of which serve as interesting directions for further related research.

Chapter 3 considers a stochastic bilevel model that simultaneously determines a supplier's trade promotion policy and a retailer's operations decisions. We presented a procedure which transforms the stochastic bilevel model to a deterministic single-level problem in the form of a generalized bilinear programming problem. We provided an exact solution method for solving this GBP problem, and compared this approach with Al-Khayyal's approach and a widely-used heuristic method for solving GBPs. Based on

our numerical study, our exact algorithm has proven to be quite efficient. In addition to the findings in Chapter 2, we again provided numerical evidence that increased supplier profit and increased system profit can coexist with the bullwhip effect as a result of price promotions if: (i) the supplier accounts for the retailer's reactions when making promotion decisions; (ii) there is a sufficient number of additional consumers who are attracted by the discounted price; and (iii) the pass-through rate is set judiciously by the retailer. This chapter builds a foundation for future research. For example, we may consider the retailer's pass-through rate as a decision variable. We are also interested in the promotion design problem when the effectiveness of a promotion depends on the supplier's and retailer's previous decisions. Finally, solving large-scale problem instances would likely necessitate the application of heuristic solution methods, wherein the methods we have proposed may be very useful for providing bounds on optimal solutions.

Chapter 4 considers a chance-constrained two stage model that determines a clicks-and-mortar retailer's assortment and pricing decisions in two selling channels. We adopted a well-developed combined sample average approximation algorithm for solving this CCTS model approximately, and our numerical results show the approximate solution obtained from the SAA algorithm will converge to the real optimal solution as the sample size increases. In addition to the SAA alogrithm, we developed our own greedy and local search heuristic, which can reduce the problem solving time significantly for large size problem instances without compromising the optimality performance significantly. This chapter built a foundation for future research on assortment and promotion planning in multiple selling channels. For example, it is hard to use the MNL model in the case where customer substitution behavior is dynamic, but the exogenous demand (EXD) model can be easily extended to include dynamic stockout substitutions. As a result, future study may consider using the EXD model in place of the MNL model to design product assortment and promotions in a similar

problem setting. We may also consider the distribution structure in which the virtual channel inventory is available to the local customers which allows a local stockout to be fulfilled from the online warehouse.

APPENDIX A
PROOF THAT RETAILER'S PROBLEM IS A CONVEX PROGRAM FOR FIXED L AND δ

We first assume that the impulsive buying rate is at least as great as the forward buying rate, and that the increase in demand rate due to forward buying is no more than four times the normal demand rate. Note that for fixed values of l and δ, D_1 is fixed and the constraints are linear in the remaining variables. The Hessian matrix of the objective function is as follows:

$$\bigtriangledown^2 PR_1(Q_1, q_1, m) = \frac{i(c-d)}{2LD_1} \begin{vmatrix} -2 & \frac{-\delta\tilde{r}_f}{D_0^*} & 0 \\ \frac{-\delta\tilde{r}_f}{D_0^*} & \frac{-2\delta(\tilde{r}_i+r_0)}{D_0^*} & 0 \\ 0 & 0 & 0 \end{vmatrix} \tag{A–1}$$

This Hessian matrix is negative semi-definite if $4\delta D_0(\tilde{r}_i + r_0) \geq \delta^2\tilde{r}_f^2$, and this condition can be written as $4D_0 \geq \delta\frac{\tilde{r}_f^2}{\tilde{r}_i+r_0}$. Next note that if $\tilde{r}_i + r_0 \geq \tilde{r}_f$, then we have $\delta\frac{\tilde{r}_f^2}{\tilde{r}_i+r_0} \leq \delta\tilde{r}_f$. The term $\delta\tilde{r}_f$ corresponds to the total forward buy from brand loyals during the promotion period. Thus, we have that the objective function is concave, provided that the condition $4D_0 \geq \delta\tilde{r}_f$ holds. That is, as long as the amount of the forward buy is no more than four times the non-discounted demand rate, our objective function is concave. The concavity of the objective and the linearity of the constraints when l and δ are fixed imply that under the conditions stated, the retailer's problem is a convex program.

APPENDIX B
PROOF OF SUFFICIENT OPTIMALITY CONDITION FOR CONVEX MIXED INTEGER PROGRAM WITH ONE INTEGER VARIABLE

Theorem B.1. *Let $f(x_1, ..., x_n, y)$ be a concave and continuously differentiable function over $S \times \mathbb{R}$, where S is an open convex subset of $\mathbb{R}^n$. Let $(x_1^*, ..., x_n^*, y^*)$ be a global optimal solution of the problem max$\{f(x_1, ..., x_n, y) | \ (x_1, ..., x_n, y) \in S \times \mathbb{R}\}$. Then we claim either $(\underline{x}_1, ..., \underline{x}_n, \lfloor y^* \rfloor) = argmax_{(x_1,...,x_n)\in S}\{f(x_1, ..., x_n, \lfloor y^* \rfloor\}$ or $(\bar{x}_1, ..., \bar{x}_n, \lceil y^* \rceil) = argmax_{(x_1,...,x_n)\in S}\{f(x_1, ..., x_n, \lceil y^* \rceil\}$ is the optimal solution of the problem max$\{f(x_1, ..., x_n, y) | \ (x_1, ..., x_n, y) \in S \times \mathbb{Z}\}$.*

Proof. First we want to show $f(\underline{x}_1, ..., \underline{x}_n, \lfloor y^* \rfloor)$ is the optimal solution of the problem

$$\max\{f(x_1, ..., x_n, y) | \ y \leq \lfloor y^* \rfloor, \ (x_1, ..., x_n, y) \in S \times \mathbb{R}\}. \tag{B–1}$$

Since $f(x_1, ..., x_n, y)$ is a concave and continuously differentiable function over $S \times \mathbb{R}$, and $y \leq \lfloor y^* \rfloor$ is also concave and continuously differentiable over $S \times \mathbb{R}$, it follows that a point $(x_1', ..., x_n', y')$ is an optimal solution of (B–1) if and only if $(x_1', ..., x_n', y')$ is an KKT point, i.e., □

$$\nabla f(x_1', ..., x_n', y') + \lambda \begin{pmatrix} 0 \\ \vdots \\ 0 \\ 1 \end{pmatrix} = 0 \tag{B–2}$$

$$\lambda(y' - \lfloor y^* \rfloor) = 0 \tag{B–3}$$

$$y' \leq \lfloor y^* \rfloor \tag{B–4}$$

$$(x_1', ..., x_n', y') \in S \times \mathbb{R} \ . \tag{B–5}$$

If $y' = \lfloor y^* \rfloor$, we are done; otherwise, we have $\lambda = 0$, which implies $\nabla f(x_1', ..., x_n', y') = 0$ and $f(x_1', ..., x_n', y') = f(x_1^*, ..., x_n^*, y^*)$. Also $y' \leq \lfloor y^* \rfloor$, so there exists $\mu \in [0, 1]$ such

that $\lfloor y^* \rfloor = \mu y' + (1-\mu) y^*$. By the properties of concave functions and the definition of $(\underline{x}_1, \ldots, \underline{x}_n, \lfloor y^* \rfloor)$, we have

$$
\begin{aligned}
& f(\underline{x}_1, \ldots, \underline{x}_n, \lfloor y^* \rfloor) \\
& \geq f(\mu x_1' + (1-\mu) x_1^*, \ldots, \mu x_n' + (1-\mu) x_n^*, \lfloor y^* \rfloor) \\
& \geq \mu f(x_1', \ldots, x_n', y') + (1-\mu) f(x_1^*, \ldots, x_n^*, y^*) \\
& = \max\{f(x_1, \ldots, x_n, y) |\ (x_1, \ldots, x_n, y) \in S \times \mathbb{R}\} \\
& \geq \max\{f(x_1, \ldots, x_n, y) |\ y \leq \lfloor y^* \rfloor),\ (x_1, \ldots, x_n, y) \in S \times \mathbb{R}\}.
\end{aligned}
$$

From above, we can conclude that $(\underline{x}_1, \ldots, \underline{x}_n, \lfloor y^* \rfloor)$ is an optimal solution of (B–1), and we can further conclude that

$$
f(\underline{x}_1, \ldots, \underline{x}_n, \lfloor y^* \rfloor) = \max\{f(x_1, \ldots, x_n, y) |\ y \leq \lfloor y^* \rfloor),\ (x_1, \ldots, x_n, y) \in S \times \mathbb{Z}\} \tag{B–6}
$$

Similarly, we can show that

$$
f(\bar{x}_1, \ldots, \bar{x}_n, \lceil y^* \rceil) = \max\{f(x_1, \ldots, x_n, y) |\ y \geq \lceil y^* \rceil),\ (x_1, \ldots, x_n, y) \in S \times \mathbb{Z}\} \tag{B–7}
$$

Combining (B–6) and (B–7), the proof is complete.

APPENDIX C
DECOMPOSITION OF UNCAPACITATED MINIMUM COST FLOW PROBLEM

This section shows a proof that an uncapacitated minimum cost flow problem can be decomposed into a set of shortest path problems.

Theorem C.1. *A single-origin, N-destination uncapacitated minimum cost flow problem can be decomposed into N shortest path problems that do not depend on the demand level.*

Proof. Assume without loss generality that we have a directed network $G = (V, E)$, where $V = \{0, 1, \dots, N\}$ is the set of nodes and E is the set of arcs. Each arc $(i, j) \in E$ has an associated cost c_{ij} that denotes the cost per unit flow on that arc. We also assume that each arc has an infinite capacity, i.e., there is no upper bound on the maximum amount that can flow on any arc. The network has a unique node 0, called the source, with supply $\sum_{i=1}^{N} d_i$. For each nonsource node $i \in V$, we associate with it a demand level d_i (≥ 0). The single-origin N-destination uncapacitated minimum cost flow problem is to determine a minimum cost flow through the uncapacitated network in order to satisfy demands at the N nonsource nodes from available supplies at the source node 0. The decision variables in this problem are arc flows, and we represent the flow on an arc $(i, j) \in E$ by x_{ij}. The single-origin N-destination uncapacitated minimum cost flow problem can be formulated as follows:

$$
\begin{aligned}
\text{(UMCF)} \quad \min \quad & \sum_{(i,j)\in E} c_{ij} x_{ij} \\
\text{s.t.} \quad & \sum_{j:(0,j)\in E} x_{0j} - \sum_{j:(j,0)\in E} x_{j0} = \sum_{i=1}^{N} d_i \\
& \sum_{j:(i,j)\in E} x_{ij} - \sum_{j:(j,i)\in E} x_{ji} = -d_i && \forall i \in \{1, \dots, N\} \\
& x_{ij} \geq 0 && \forall (i,j) \in E \ .
\end{aligned}
$$

The flow decomposition theorem [5] implies that the flow x_{ij} on each arc $(i,j) \in E$ can be decomposed by destination, with x_{ij}^l representing of flow on arc (i,j) used to satisfy demand at node l. With this notation, we can reformulate the single-origin N-destination uncapacitated minimum cost flow problem as follows:

$$
\begin{aligned}
\text{(UMCF')} \quad \min \quad & \sum_{l=1}^{N} \sum_{(i,j)\in E} c_{ij} x_{ij}^l \\
\text{s.t.} \quad & \sum_{j:(0,j)\in E} x_{0j}^l - \sum_{j:(j,0)\in E} x_{j0}^l = d_l && \forall l \in \{1,\dots,N\} \\
& \sum_{j:(i,j)\in E} x_{ij}^l - \sum_{j:(j,i)\in E} x_{ji}^l = \begin{cases} 0 & \text{if } i \neq l \\ -d_l & \text{if } i = l \end{cases} && \forall i, l \in \{1,\dots,N\} \\
& x_{ij}^l \geq 0 && \forall (i,j) \in E, l \in \{1,\dots,N\} \ .
\end{aligned}
$$

By definition, we have $x_{ij} = \sum_{l=1}^{L} x_{ij}^l$. Suppose we define variables $y_{ij}^l = \frac{x_{ij}^l}{d_l}$. Using these, the single-origin N-destination uncapacitated minimum cost flow problem can be expressed as follows:

$$
\begin{aligned}
\text{(UMCF")} \quad \min \quad & \sum_{l=1}^{N} d_l \sum_{(i,j)\in E} c_{ij} y_{ij}^l \\
\text{s.t.} \quad & \sum_{j:(0,j)\in E} y_{0j}^l - \sum_{j:(j,0)\in E} y_{j0}^l = 1 && \forall l \in \{1,\dots,N\} \\
& \sum_{j:(i,j)\in E} y_{ij}^l - \sum_{j:(j,i)\in E} y_{ji}^l = \begin{cases} 0 & \text{if } i \neq l \\ -1 & \text{if } i = l \end{cases} && \forall i, l \in \{1,\dots,N\} \\
& y_{ij}^l \geq 0 && \forall (i,j) \in E, l \in \{1,\dots,N\} \ .
\end{aligned}
$$

In this above formulation, the set of feasible solutions Y can be decomposed into N sets $Y^1, Y^2, \dots, Y^N$, where

$$Y^l \equiv \left\{ y^l : \sum_{j:(i,j)\in E} y^l_{ij} - \sum_{j:(j,i)\in E} y^l_{ji} = \begin{cases} 1 & \text{if } i = 0 \\ 0 & \text{if } i \neq l \\ -1 & \text{if } i = l \end{cases} \forall (i,j) \in E \right\} .$$

By observation, the variables for each subset do not appear in any other set. Consequently, we can decompose UMCP" into N shortest path problems which do not dependent on the demand level in the following formulation:

$$\begin{aligned}
(SP^l) \ \min \ & d_l \sum_{(i,j)\in E} c_{ij} y^l_{ij} = d_l \min \sum_{(i,j)\in E} c_{ij} y^l_{ij} \\
\text{s.t.} \ & \sum_{j:(0,j)\in E} y^l_{0j} - \sum_{j:(j,0)\in E} y^l_{j0} = 1 \\
& \sum_{j:(i,j)\in E} y^l_{ij} - \sum_{j:(j,i)\in E} y^l_{ji} = \begin{cases} 0 & \text{if } i \neq l \\ -1 & \text{if } i = l \end{cases} && \forall i \in \{1, \dots, N\} \\
& y^l_{ij} \geq 0 && \forall (i,j) \in E \ .
\end{aligned}$$

After solving all of these N shortest path problems, we can obtain the solutions to the original UMCP using the following relationship:

$$x_{ij} = \sum_{l=1}^{L} x^l_{ij} = \sum_{l=1}^{L} d_l y^l_{ij} \ \forall (i,j) \in E \ .$$

□

APPENDIX D
DETAILED AL-KHAYYAL'S APPROACH

The detailed Al-Khayyal's branch-and-bound algorithm is shown below.

Algorithm 2 Al-Khayyal's Branch-and-Bound Algorithm

1: $\Omega^{(1,1)} \leftarrow \Omega$, $T_1 \leftarrow \{(1,1)\}$, $k \leftarrow 1$, $UB^{(1,1)} \leftarrow \infty$, $UB \leftarrow \infty$ and $LB \leftarrow -\infty$;
2: **while** $UB - LB > \epsilon$ **do**
3: choose an active node $(u, v) \in T_k$ and remove node (u, v) from active node set;
4: $\bar{\Lambda} \leftarrow$ an optimal solution to $LP(\Omega^{(u,v)})$;
5: $UB_k \leftarrow \Phi(\bar{\Lambda})$;
6: **if** $(\bar{x}, \bar{y}, \bar{z}, \bar{\pi})$ satisfying constraint (3–9) **then**
7: $LB_k \leftarrow \phi(\bar{x}, \bar{y}, \bar{z}, \bar{\pi})$;
8: **if** $LB_k < UB_k$ **then**
9: $(p, q) \leftarrow \arg\max_{(i,l)}\{\bar{z}_i^l \bar{x}_i^l - \bar{\eta}_i^l\}$
10: **else**
11: $(p, q) \leftarrow (0, 0)$;
12: **end if**
13: **else**
14: $LB_k \leftarrow -\infty$;
15: $(p, q) \leftarrow \arg\max_{(i,l)} \left\{\max\left[\bar{c}_i^l x_i^l + \sum_{j \neq i} \bar{t}_{ij}^l y_{ij}^l - \bar{z}_i^l \bar{x}_i^l - \bar{\zeta}_i^l, - \sum_{\tau=l}^{L} (d_i^l \bar{\pi}_i^\tau + \alpha_i^l \gamma_i^l \bar{z}_i^l \bar{\pi}_i^\tau + \bar{\nu}_i^{l\tau})\right]\right\}$;
16: **end if**
17: **if** $(p, q) \neq (0, 0)$ **then**
18: $\Omega^{(k+1,1)} \leftarrow \Omega^{(u,v)} \cap \{ZL_p^q \leq z_p^q \leq \bar{z}_p^q\}$;
19: $\Omega^{(k+1,2)} \leftarrow \Omega^{(u,v)} \cap \{\bar{z}_p^q \leq z_p^q \leq ZU_p^q\}$;
20: $UB^{(k+1,i)} \leftarrow UB_k$, $\forall i = 1, 2$;
21: **end if**
22: **if** $LB_k > LB$ **then**
23: $LB \leftarrow LB_k$ and $(x^*, y^*, z^*, \pi^*) \leftarrow (\bar{x}, \bar{y}, \bar{z}, \bar{\pi})$;
24: $T_k \leftarrow T_k - \{(m, n) \in T_k : UB^{(m,n)} \leq LB\}$;
25: $T_{k+1} \leftarrow T_k \cup \{(k+1, 1), (k+1, 2)\}$;
26: $UB = \max\{UB^{(m,n)}, \forall (m, n) \in T_{k+1}\}$;
27: **end if**
28: $k \leftarrow k + 1$;
29: **end while**

After initializing the branch-and-bound tree in line 1, the algorithm repeats the *Main Step* in the **while** loop in lines 2 - 29 until the optimality gap is less than a predetermined value ϵ. At the start of each iteration k of the **while** loop, a node (u, v) is selected and removed from tree T_k in line 3. In lines 4 and 5, the linear program $LP(\Omega^{(u,v)})$ is solved to obtain the solution $\bar{\Lambda}$ and the corresponding optimal value $\Phi(\bar{\Lambda})$ is assigned to UB_k.

In lines 6 - 16, the current iteration lower bound LB_k is obtained and the partitioning index (p, q) is found by checking the optimality and the feasibility of the solution $\bar{\Lambda}$. After finding (p, q), partition the region $\Omega^{(u,v)}$ into two mutually exclusive and exhaustive subregions in lines 18 - 21. At the end of the **while** loop, the branch-and-bound tree is updated by three types of operations in lines 22 - 27. Included here are

1. update the lower bound of the problem if $LB < LB_k$ as follows:

$$LB \leftarrow LB_k \text{ and } (x^*, y^*, z^*, \pi^*) \leftarrow (\bar{x}, \bar{y}, \bar{z}, \bar{\pi});$$

2. update the node set at iteration k as

$$T_k \leftarrow T_k - \{(m, n) \in T_k : UB^{(m,n)} \leq LB\};$$

3. update the upper bound of the problem as

$$UB = \max\{UB^{(m,n)}, \forall (m, n) \in T_{k+1}\}.$$

APPENDIX E
NUMERICAL RESULTS FOR CHAPTER 4

Table E-1. Computational results I - instance 2.

$(M, \bar{L})$	L	LB	UB	Gap = $\frac{LB}{UB} \times 100$ (%)	Time(s)
(25,1000)	20	29664	29722	0.19	147
	50	29674	29703	0.09	209
	100	29674	29689	0.05	436
(25,2000)	20	29673	29708	0.11	291
	50	29674	29697	0.07	432
	100	29674	29674	0.00	949
(50,1000)	20	29669	29741	0.24	254
	50	29670	29678	0.02	319
	100	29674	29676	0.01	828
(50,2000)	20	29666	29721	0.18	493
	50	29669	29691	0.07	630
	100	29674	29674	0.00	1106

Table E-2. Computational results I - instance 3.

$(M, \bar{L})$	L	LB	UB	Gap = $\frac{LB}{UB} \times 100$ (%)	Time(s)
(25,1000)	20	32678	32711	0.10	138
	50	32682	32702	0.05	206
	100	32681	32701	0.05	406
(25,2000)	20	29667	32727	0.18	264
	50	32670	32704	0.10	386
	100	32678	29684	0.02	745
(50,1000)	20	32673	32733	0.18	254
	50	32671	32701	0.09	319
	100	32670	32683	0.03	520
(50,2000)	20	32666	32722	0.17	487
	50	29676	32690	0.04	629
	100	32674	32683	0.02	1004

Table E-3. Computational results I - instance 4.

$(M, \bar{L})$	L	LB	UB	Gap = $\frac{LB}{UB} \times 100$ (%)	Time(s)
(25,1000)	20	32492	32591	0.30	132
	50	32491	32529	0.11	183
	100	32502	32513	0.03	325
(25,2000)	20	32497	32557	0.18	268
	50	32501	32521	0.06	362
	100	32498	32501	0.01	742
(50,1000)	20	32494	32555	0.18	239
	50	32493	32516	0.07	290
	100	32497	32502	0.02	479
(50,2000)	20	32492	32568	0.23	473
	50	32498	32529	0.09	571
	100	32496	32499	0.01	857

Table E-4. Computational results I - instance 4.

$(M, \bar{L})$	L	LB	UB	Gap = $\frac{LB}{UB} \times 100$ (%)	Time(s)
(25,1000)	20	37866	39412	0.30	132
	50	32491	32529	0.11	183
	100	32502	32513	0.03	325
(25,2000)	20	32497	32557	0.18	268
	50	32501	32521	0.06	362
	100	32498	32501	0.01	742
(50,1000)	20	32494	32555	0.18	239
	50	32493	32516	0.07	290
	100	32497	32502	0.02	479
(50,2000)	20	32492	32568	0.23	473
	50	32498	32529	0.09	571
	100	32496	32499	0.01	857

REFERENCES

[1] "Pay-for-performance trade promotions can ease friction between manufacturers and retailers." *Knowledge@Wharton* (2001).

[2] Agatz, N. A., Fleischmann, M., and Van Nunen, J. A. "E-fulfillment and multi-channel distribution–A review." *European Journal of Operational Research* 187 (2008).2: 339–356.

[3] Ahmed, S. and Shapiro, A. "Solving chance-constrained stochastic programs via sampling and integer programming." *Tutorials in Operations Research,(Z.-L. Chen and S. Raghavan, eds.), INFORMS* (2008).

[4] Ahmed, S., Shapiro, A., and Shapiro, E. "The sample average approximation method for stochastic programs with integer recourse." *SIAM Journal of Optimization* 12 (2002): 479–502.

[5] Ahuja, R. K., Magnanti, T. L., and Orlin, J. B. *Network Flows: Theory, Algorithms, and Applications*. Upper Saddle River, NJ: Prentice-Hall, 1993.

[6] Ailawadi, K., Farris, P. W., and Shames, E. "Trade promotion: Essential to selling through resellers." *Sloan Management Review* 41 (1999).1: 83–92.

[7] Al-Khayyal, F. A. "Generalized bilinear programming: Part I. models, applications and linear programming relaxation." *European Journal of Operational Research* 60 (1992).3: 306–314.

[8] Al-Khayyal, F. A., C.Larsen, and Voorhis, T. Van. "A relaxation method for nonconvex quadratically constrained quadratic programs." *Journal of Global Optimization* 6 (1995).3: 215–230.

[9] Al-Khayyal, F. A. and Falk, J. E. "Jointly constrained biconvex programming." *Mathematics of Operations Research* 8 (1983).2: 273–286.

[10] Anderson, E. E. and Amato, H. N. "A mathematical model for simultaneously determining the optimal brand-collection and display-area allocation." *Operations Research* 22 (1974).1: 13–21.

[11] Anderson, S. P., Palma, A. De, and Thisse, J. F. *Discrete choice theory of product differentiation*. Cambridge, MA: The MIT Press, 1992.

[12] Arcelus, F. J. and Srinivasan, G. "Ordering policies under one time only discount and price sensitive demand." *IIE Transactions* 30 (1998).11: 1057–1064.

[13] Ardalan, A. "Optimal ordering policies in response to a sale." *IIE Transactions* 20 (1988).3: 292–294.

[14] ———. “Combined optimal price and optimal inventory replenishment policies when a sale results in increase in demand.” *Computers & Operations Research* 18 (1991).8: 721–730.

[15] ———. “Optimal prices and order quantities when temporary price discounts result in increase in demand.” *European Journal of Operational Research* 72 (1994).1: 52–61.

[16] Ayanso, A., Diaby, M., and Nair, S. K. “Inventory rationing via drop-shipping in Internet retailing: A sensitivity analysis.” *European Journal of Operational Research* 171 (2006).1: 135–152.

[17] Bae, S. H., Nam, S., and Kim, S. H. “Empirical Investigation of Consumers’ Impulse Purchases from Television Home Shopping Channels: A Case of Order Cancellation Behavior.” *Working paper* (2011).

[18] Bailey, J. P. and Rabinovich, E. “Internet book retailing and supply chain management: an analytical study of inventory location speculation and postponement.” *Transportation Research Part E: Logistics and Transportation Review* 41 (2005).3: 159–177.

[19] Bell, D. R. and Drèze, X. “Changing the channel: A better way to do trade promotions.” *MIT Sloan management review* 43 (2002).2: 42–49.

[20] Bell, D. R. and Lattin, J. M. “Shopping behavior and consumer preference for store price format: Why” large basket” shoppers prefer EDLP.” *Marketing Science* (1998): 66–88.

[21] Ben-Ayed, O. “Bilevel linear programming.” *Computers & Operations Research* 20 (1993).5: 485–501.

[22] Berry, J. “Diverting: Trade diverting is becoming ever more widespread. As it grows, so does mistrust between retailer and manufacturer.” *Adweek’s Marketing Week* 33 (1992).20.

[23] Besanko, D., Dubé, J. P., and Gupta, S. “Own-brand and cross-brand retail pass-through.” *Marketing Science* 24 (2005).1: 123–137.

[24] Blattberg, R. C., Briesch, R., and Fox, E. J. “How promotions work.” *Marketing Science* 14 (1995).3: 122–132.

[25] Blattberg, R. C., Eppen, G. D., and Lieberman, J. “A theoretical and empirical evaluation of price deals for consumer nondurables.” *The Journal of Marketing* 45 (1981).1: 116–129.

[26] Blattberg, R. C. and Neslin, S. A. “Sales promotion: The long and the short of it.” *Marketing Letters* 1 (1989).1: 81–97.

[27] Borin, N. and Farris, P. “A sensitivity analysis of retailer shelf management models.” *Journal of Retailing* 71 (1995).2: 153–171.

[28] Bronnenberg, B. J., Dhar, S. K., and Dubé, J. P. “Market structure and the geographic distribution of brand shares in consumer package goods industries.” *University of Chicago Mimeograph* (2005).

[29] ———. “Consumer packaged goods in the United States: National brands, local branding.” *Journal of Marketing Research* 44 (2007).1: 4–13.

[30] Bront, J. J. M., Méndez-Díaz, I., and Vulcano, G. “A column generation algorithm for choice-based network revenue management.” *Operations Research* 57 (2009).3: 769–784.

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

[32] Cachon, G. P., Terwiesch, C., and Xu, Y. “Retail assortment planning in the presence of consumer search.” *Manufacturing & Service Operations Management* 7 (2005).4: 330–346.

[33] Cattani, K. D. and Souza, G. C. “Inventory rationing and shipment flexibility alternatives for direct market firms.” *Production and Operations Management* 11 (2002).4: 441–457.

[34] Chandon, P., Wansink, B., and Laurent, G. “A benefit congruency framework of sales promotion effectiveness.” *The Journal of Marketing* 64 (2000).4: 65–81.

[35] Chen, F., Drezner, Z., Ryan, J. K., and Simchi-Levi, D. “Quantifying the bullwhip effect in a simple supply chain: The impact of forecasting, lead times, and information.” *Management Science* 46 (2000).3: 436–443.

[36] Chen, F., Ryan, J. K., and Simchi-Levi, D. “The impact of exponential smoothing forecasts on the bullwhip effect.” *Naval Research Logistics* 47 (2000).4: 269–286.

[37] Chen, K. D. and Hausman, W. H. “Technical note: Mathematical properties of the optimal product line selection problem using choice-based conjoint analysis.” *Management Science* 46 (2000).2: 327–332.

[38] Cheng, F. and Sethi, S. P. “A periodic review inventory model with demand influenced by promotion decisions.” *Management Science* 45 (1999).11: 1510–1523.

[39] Corstjens, M. and Doyle, P. “A model for optimizing retail space allocations.” *Management Science* 27 (1981).7: 822–833.

[40] Ding, Q., Kouvelis, P., and Milner, J. M. “Dynamic pricing through discounts for optimizing multiple-class demand fulfillment.” *Operations Research* 54 (2006).1: 169–183.

[41] Dong, L., Kouvelis, P., and Tian, Z. "Dynamic pricing and inventory control of substitute products." *Manufacturing & Service Operations Management* 11 (2009).2: 317–339.

[42] Drèze, X. and Bell, D. R. "Creating win-win trade promotions: Theory and empirical analysis of scan-back trade deals." *Marketing Science* 22 (2003).1: 16–39.

[43] Dukes, A. J., Geylani, T., and Srinivasan, K. "Strategic assortment reduction by a dominant retailer." *Marketing Science* 28 (2009).2: 309–319.

[44] Eliashberg, J. and Steinberg, R. "Marketing-production decisions in an industrial channel of distribution." *Management Science* 33 (1987).8: 981–1000.

[45] Engineering, User Interface. "What causes customers to buy on impulse." (2002).

[46] Ernst, R. and Pyke, D. F. "Optimal base stock policies and truck capacity in a two-echelon system." *Naval Research Logistics* 40 (1993).7: 879–903.

[47] Forrester, J. W. and Wright, J. *Industrial Dynamics*. MIT press Cambridge, MA, 1961.

[48] Fransoo, J. C. and Wouters, M. J. F. "Measuring the bullwhip effect in the supply chain." *Supply Chain Management: An International Journal* 5 (2000).2: 78–89.

[49] Gallego, G. and Wang, R. "Price optimization and competition for multi-products under the nested logit model with product-differentiated price coefficients." *Working paper* (2011).

[50] Gavirneni, S. "Price fluctuations, information sharing, and supply chain performance." *European Journal of Operational Research* 174 (2006).3: 1651–1663.

[51] Gedenk, K., A, S. Neslin, and Ailawadi, K. "Sales Promotion." *Retailing in the 21st Century* (2006): 345–359.

[52] Gómez, M. I., Rao, V. R., and McLaughlin, E. W. "Empirical analysis of budget and allocation of trade promotions in the US supermarket industry." *Journal of Marketing Research* 44 (2007).3: 410–424.

[53] Goyal, S. K. "An integrated inventory model for a single product system." *Operational Research Quarterly* 28 (1977).3: 539–545.

[54] Gupta, S. "Impact of sales promotions on when, what, and how much to buy." *Journal of Marketing Research* (1988): 342–355.

[55] Hansen, P. and Heinsbroek, H. "Product selection and space allocation in supermarkets." *European Journal of Operational Research* 3 (1979).6: 474–484.

[56] Hanson, W. and Martin, K. "Optimizing multinomial logit profit functions." *Management Science* 42 (1996).7: 992–1003.

[57] Hoch, S. J., Dreze, X., and Purk, M. E. "EDLP, Hi-Lo, and margin arithmetic." *The Journal of Marketing* (1994): 16–27.

[58] Hwang, H., Choi, B., and Lee, M. "A model for shelf space allocation and inventory control considering location and inventory level effects on demand." *International Journal of Production Economics* 97 (2005).2: 185–195.

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

[60] Kogan, K. and Herbon, A. "A supply chain under limited-time promotion: The effect of customer sensitivity." *European Journal of Operational Research* 188 (2008).1: 273–292.

[61] Kök, A. G., Fisher, M. L., and Vaidyanathan, R. "Assortment planning: Review of literature and industry practice." *Retail Supply Chain Management* (2009): 1–55.

[62] Kopalle, P. K., Mela, C. F., and Marsh, L. "The dynamic effect of discounting on sales: Empirical analysis and normative pricing implications." *Marketing Science* 18 (1999).3: 317–332.

[63] Kril, J. "Statistics and Facts about Online Shopping." *http://www.statista.com/topics/871/online-shopping/* (2012).

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

[65] ———. "The Bullwhip Effect In Supply Chains1." *Sloan Management Review* 38 (1997).3: 93–102.

[66] Li, H. and Huh, W. T. "Pricing multiple products with the multinomial logit and nested logit models: Concavity and implications." *Manufacturing & Service Operations Management* 13 (2011).4: 549–563.

[67] Liebmann, W. "Online retailers moving into offline shopping fast." (2013).

URL http://www.forbes.com/sites/wendyliebmann/2013/05/30/online-goes-offline-fast/

[68] Liu, Q. and van Ryzin, G. "On the Choice-Based Linear Programming Model for Network Revenue Management." *Manufacturing & Service Operations Management* 10 (2008).2: 288–310.

[69] Luedtke, J., Ahmed, S., and Nemhauser, G. L. "An integer programming approach for linear programs with probabilistic constraints." *Mathematical Programming* 122 (2010).2: 247–272.

[70] Mahajan, S. and Van Ryzin, G. "Stocking retail assortments under dynamic consumer substitution." *Operations Research* 49 (2001).3: 334–351.

[71] Miller, C. M., Smith, S. A., McIntyre, S. H., and Achabal, D. D. "Optimizing and evaluating retail assortments for infrequently purchased products." *Journal of Retailing* 86 (2010).2: 159–171.

[72] Miller, R. "Does everyone have a price." *Marketing* 30 (1997).3.

[73] Mittelstaedt, R. A. and Stassen, R. E. "Shopping behavior and retail merchandising strategies." *Journal of Business Research* 21 (1990).3: 243–258.

[74] Muu, L. D. and Quy, N. V. "A global optimization method for solving convex quadratic bilevel programming problems." *Journal of Global Optimization* 26 (2003).2: 199–219.

[75] MWPVL International Inc. "The grocery distribution center network in North America." (2010).

[76] Naddor, E. *Inventory Systems*. Wiley New York, 1966.

[77] Neslin, S. A., Powell, S. G., and Stone, L. S. "The effects of retailer and consumer response on optimal manufacturer advertising and trade promotion strategies." *Management Science* 41 (1995).5: 749–766.

[78] Netessine, S. and Rudi, N. "Supply chain choice on the internet." *Management Science* 52 (2006).6: 844–864.

[79] Petrik, M. and Zilberstein, S. "Robust Approximate Bilinear Programming for Value Function Approximation." *Journal of Machine Learning Research* 12 (2011): 3027–3063.

[80] Petruzzi, N. C. and Dada, M. "Pricing and the newsvendor problem: A review with extensions." *Operations Research* 47 (1999).2: 183–194.

[81] Potter, A. and Disney, S. M. "Bullwhip and batching: An exploration." *International Journal of Production Economics* 104 (2006).2: 408–418.

[82] Randall, T., Netessine, S., and Rudi, N. "Should you take the virtual fulfillment path?" *Supply Chain Management Review* 6 (2002).6: 54 – 58.

[83] ———. "An empirical examination of the decision to invest in fulfillment capabilities: A study of Internet retailers." *Management Science* 52 (2006).4: 567–580.

[84] Sherali, H. D. and Alameddine, A. "A new reformulation-linearization technique for bilinear programming problems." *Journal of Global Optimization* 2 (1992).4: 379–410.

[85] Sherali, H. D. and Shetty, C. M. "A finitely convergent algorithm for bilinear programming problems using polar cuts and disjunctive face cuts." *Mathematical Programming* 19 (1980).1: 14–31.

[86] Singh, P., Groenevelt, H., and Rudi, N. "Product variety and supply chain structures." *Working paper* (2005).

[87] Smith, S. A. and Agrawal, N. "Management of multi-item retail inventory systems with demand substitution." *Operations Research* (2000): 50–64.

[88] Sogomonian, A. G. and Tang, C. S. "A modeling framework for coordinating promotion and production decisions within a firm." *Management Science* 39 (1993).2: 191–203.

[89] Song, J. and Xue, Z. "Demand management and inventory control for substitutable products." *Working paper* (2007).

[90] Sterman, J. D. "Misperceptions of feedback in dynamic decision making." *Organizational Behavior and Human Decision Processes* 43 (1989).3: 301–335.

[91] Taylor, S. G. and Bradley, C. E. "Optimal ordering strategies for announced price increases." *Operations Research* 33 (1985).2: 312–325.

[92] Thowsen, G. T. "A dynamic, nonstationary inventory problem for a price/quantity setting firm." *Naval Research Logistics Quarterly* 22 (1975).3: 461–476.

[93] Topaloglu, H. "Joint Stocking and Product Offer Decisions Under the Multinomial Logit Model." *Production and Operations Management* 22 (2013).5: 1182–1199.

[94] Urban, T. L. "An inventory-theoretic approach to product assortment and shelf-space allocation." *Journal of Retailing* 74 (1998).1: 15–35.

[95] Van Ryzin, G. and Mahajan, S. "On the relationship between inventory costs and variety benefits in retail assortments." *Management Science* (1999): 1496–1509.

[96] Vicente, L. N., Calamai, P. H., and Júdice, J. J. "Generation of disjointly constrained bilinear programming test problems." *Computational Optimization and Applications* 1 (1992).3: 299–306.

[97] Wang, Q., Guan, Y., and Wang, J. "A Chance-Constrained Two-Stage Stochastic Program for Unit Commitment With Uncertain Wind Power Output." *IEEE Transactions on Power Systems* (2011).99: 1–1.

[98] Wen, U. P. and Hsu, S. T. "Linear bi-level programming problems–a review." *Journal of the Operational Research Society* 42 (1991).2: 125–133.

[99] Wu, T. H. "A note on a global approach for general 0–1 fractional programming." *European Journal of Operational Research* 101 (1997).1: 220–223.

[100] Xie, J. and Wei, J. “Coordinating advertising and pricing in a manufacturer–retailer channel.” *European Journal of Operational Research* 197 (2009).2: 785–791.

[101] Zufryden, F. S. “A dynamic programming approach for product selection and supermarket shelf-space allocation.” *Journal of the Operational Research Society* (1986): 413–422.

BIOGRAPHICAL SKETCH

Yiqiang Su was born in Beijing in 1987. Yiqiang is the oldest son of Changcheng Su and Bingyi Ji. Yiqiang earned his bachelor's degree in industrial engineering and engineering management from the Hong Kong University of Science and Technology (HKUST) in 2009, and he was also the recipient of the Academic Achievement Award which is the highest academic honor bestowed by the University on undergraduates upon graduation. Yiqiang joined the Department of Industrial and Systems Engineering at the University of Florida (UF) in August 2009, starting his doctoral study under the guidance of Dr. Joseph Geunes. He received his Ph.D. from the University of Florida in the fall of 2013. Following graduation, he joins the BNSF Railway Company as a senior operations research specialist.