Three Essays in Operations Management

by

Ngai-Hang Zachary Leung

Submitted to the Sloan School of Management
in partial fulfillment of the requirements for the degree of

Doctor of Philosophy in Operations Research

at the

MASSACHUSETTS INSTITUTE OF TECHNOLOGY

September 2014

Signature redacted

Author ....
Sloan School of Management
August 15, 2014

Signature redacted

Certified by ....
Georgia Perakis
William F. Pounds Professor of Operations Research
Thesis Supervisor

Signature redacted

Accepted by ....
Dimitris Bertsimas
Boeing Leaders for Global Operations Professor
Codirector, Operations Research Center

Three Essays in Operations Management

by

Ngai-Hang Zachary Leung

Submitted to the Sloan School of Management
on August 15, 2014, in partial fulfillment of the
requirements for the degree of
Doctor of Philosophy in Operations Research

Abstract

The thesis applies optimization theory to three problems in operations management.

In the first part of the thesis, we investigate the impact of inventory control on the availability of drugs to patients at public health facilities in Zambia. We present consistent empirical data and simulation results showing that, because of its failure to properly anticipate seasonal variations in demand and supply lead-times, this system leads to predictable patient-level stock-outs even when there is ample inventory available in the central warehouse. Secondly, we propose an alternative inventory control system relying on mobile devices and mathematical optimization, and present results from a validated simulation model suggesting that its implementation would lead to a substantial improvement of patient access to drugs relative to the current system.

In the second part of the thesis, we investigate the impact of returning customers on pricing for fashion Internet retailers. Our analysis of clickstream data from an online fashion retailer shows that a significant proportion of sales is due to returning customers, i.e. customers who first visit an item at a particular price, but purchase the item in a later visit at a lower price. We propose a markdown pricing model that explicitly incorporates returning customers. We propose a model for quantifying the value of the returning pricing model relative to a pricing model that does not distinguish between first-time and returning customers, and determine the value of returning pricing both exactly and through developing bounds. Based on real data from a fashion Internet retailer, we estimate the parameters of the returning demand model and determine the value of the returning pricing model.

Lastly, we study the promotion optimization problem faced by grocery retailers, i.e. deciding which items to promote and at what price. Our formulation includes several business rules that arise in practice. We build demand models from data in order to capture the stockpiling behavior through dependence on past prices. This gives rise to a hard problem. For general additive and multiplicative demand structures, we propose efficient LP based methods, show theoretical performance guarantees and validate our results using real data.

Thesis Supervisor: Georgia Perakis
Title: William F. Pounds Professor of Operations Research

Acknowledgments

I would first like to thank my adviser Georgia Perakis, without whom this thesis could not have been completed. Thank you Georgia for your encouragement and support, especially during times when I felt discouraged. I really appreciate how you made time to meet me weekly, and during certain busy seasons, even daily. I also learned a lot from you not just about research, but from your perspective and advice regarding academic life.

I would also like to thank Jérémie Gallien who was my adviser for the Zambia medical supply chapter of the thesis. Thank you for your funding during the first two years of my doctoral studies. This includes funding for an eye-opening field visit to Zambia that we went on together with Prashant, which is one of my best memories of MIT. I appreciate the unique research perspective and your desire to make a real-world impact through your research.

I enjoyed learning from Prashant Yadav about the interesting and important challenges faced by many countries with regard to public health. Your knowledge of the challenges on the ground was invaluable in our Zambia medical supply chain project.

I feel very fortunate to collaborated with Pavithra Harsha. I benefitied from her industry experience, and her help in working through some of the difficult technical proofs in this thesis.

It was a good experience for me to work with Maxime Cohen, and I benefited from his help in checking and polishing some of the proofs in this thesis, as well as his help in writing our paper.

Jim Orlin hired me as a teaching assistant for his class 15.058 "Optimization Methods in Management Science" which was my first teaching experience at MIT. I enjoyed Jim's lectures and his unique teaching style, as well as working with fellow TA Giacomo Nannicini.

I had two opportunities to work with Charlie Fine as a teaching assistant, for the classes 15.761 "Introduction to Operations Management" and 15.778 "Management of

Supply Networks for Products and Services." I have benefited from observing his teaching and from interacting with MBA and Sloan Fellow students.

I feel very priviliged to have the opportunity to have interacted with so many creative and wise academic mentors at MIT, and many smart, creative and fun classmates at the Operations Research Center. It is my hope that I can "pay it forward" by showing students the same concern that my mentors showed me and to passing on their words of wisdom.

Life in Cambridge would have been very lonely and dull without the wonderful friends whom I spent time with here. In my first two years at Cambridge, it was a magical experience to live in Tang Hall together with the so-called "Tang small group" which was comprised of Yong, Samuel, George, Marcus, Stephanie, Michelle, Heather, Emily. I also treasure the weekly meetings we had with the so-called "Fellowship of the Ringed" group: Joe and Kayleen, Eric and Victoria, Kwabena and Lisseth. Finally, the couples and families whom we met every two weeks: Michael and Florence Yee, Mark and Vivian Shu, Tim and Ee San Ng, and Ping, Anna and Elizabeth He.

I am also thankful for the friends with whom I could share anti-stress exercise activies to stay sane: the ORC soccer team, my tennis friends (Yong, Mike Fong, Peter Oberhofer) and basketball friends (Peng, Yukkee, Ming, Dan, Justin, Kenan).

Lastly, I am thankful to God for because being at MIT is a dream come true. These last six years are extra-sweet because during this time, I married my wonderful wife, and we saw the birth of our two healthy and happy daughters. Without the love and support of my parents and Gloria's parents, my kids, and especially my long-suffering wife, I would not have been able to complete this thesis.

to God be the glory, great things He has done!

Contents

List of Figures

List of Tables

Chapter 1

Introduction

Operations research emerged as a formal discipline during World War II, due to the efforts of military planners to use analytical methods to make better decisions in their military operations. The early pioneers of operations research applied techniques including mathematical modeling, statistical analysis and mathematical optimization to a diverse range of practical problems, including optimizing the size of warship convoys to protect merchant ships against attacks by German U-boats, and determining which color was best to camouflage anti-submarine aircraft. The author's aspiration is that this thesis embodies the creative and enterprising spirit of the forefathers of operations research.

In this thesis, we apply operations research techniques to address three interesting and important practical operational problems.

In Chapter 2, we investigate the impact of the inventory control policies on the availability of essential medical drugs to patients at public health facilities in Zambia. We build a validated simulation model of the Zambia supply chain and propose an optimization-based inventory control policy that achieves a higher level of availability of drugs at health facilities compared to the existing inventory control policy.

In Chapter 3, we formulate a price optimization model for fashion Internet retailers. Our model leverages returning behavior of customers which the retailer can observe via clickstream data in order to make better pricing decisions, relative to a myopic pricing model that is based on aggregate sales data and ignores customers' returning behavior. Our numerical experiments based on real data suggest that our model would increase the retailer's revenue relative to the myopic pricing model.

In Chapter 4, we formulate an promotion optimization model for grocery retailers. Our model incorporates features that are important in practice: namely, business rules that constrain the promotion schedule, as well as demand functions with promotion fatigue effects. Because the promotion optimization problem is nonconvex, we propose an efficient LP-based approximation for which we derive theoretical guarantees. Finally, we validate our results using real data.

Finally, I present my concluding remarks in Chapter 5.

Chapter 2

Inventory Management for Essential Drugs in Sub-Saharan Africa

The following chapter was written in collaboration with Jérémie Gallien and Prashant Yadav.

2.1 Introduction

Well functioning health systems rely on the uninterrupted and continuous availability of medicines, vaccines, diagnostics and other medical supplies [Management Sciences for Health, 1997]. Unfortunately, the average service level of drugs at public health facilities in sampled low income countries has recently been found lower than 25% [Cameron et al., 2009] with the availability of malaria drugs being a specific concern [Zurovac et al., 2007, 2008]. Drug stockouts have negative impacts on public health, including widespread treatment discontinuation possibly leading to death and risk of increased resistance to drug in the overall population [World Health Organization, 2004, Levine et al., 2008, Pasquet et al., 2010]. The reasons for drug stockouts in resource-limited countries are varied and interconnected, including for example funding shortages, errors in estimation of drug procurement needs and limited available information for inventory management [Kangwana et al., 2009].

Zambia is fairly representative of the disease burden, pharmaceutical distribution and drug stockout situation in sub-Saharan Africa: its under five infant mortality rate is 141 compared to the sub-Saharan average of 129 (World Bank statistics), with malaria acting as a key driver of child mortality; drugs are received and stored in a central warehouse and supplied first to 72 district medical stores then to health facilities in the face of substantial transportation, staffing and infrastructure challenges [Yadav et al., 2011, 2012]; assessments conducted in recent years reported high rates of drug stockouts at the health facility level [Picazo and Zhao, 2009, Friedman et al., 2011]. However, the Zambian Ministry of Health (MoH) and its partners have invested significant resources in the public sector supply chain for essential medicines in recent years, resulting in fewer procurement delays and better distribution at the first tier of the distribution system [Yadav et al., 2012]. In addition, the MoH and several key partners[1] conducted in 2009 a landmark randomized pilot experiment to help determine the appropriate number of inventory holding tiers and loci of order quantity decisions [Friedman et al., 2011].

Leveraging data collected during this pilot, the present paper specifically focuses on inventory control, i.e. the set of rules and processes used to determine the quantity of each drug to be sent to each facility at each point in time, i.e., the supply chain's nervous system. In particular, it shows that a significant improvement of patient access to drugs can be achieved through changes in the inventory control policy currently implemented in Zambia which is similar to the policies and guidelines used in many low income country medicine supply chains. (For an example of such guidelines see USAID | DELIVER [2011a].) After a literature review (in §2.1) and further background on Zambia's pharmaceutical distribution system (in §2.2), it first offers (in §2.5) an original performance analysis of Zambia's existing inventory control system relying on a new dataset constructed by digitizing stock reporting cards in rural health clinics of several districts. Secondly, this paper develops (in §2.4) a detailed proposal for an alternative inventory control system relying on (i) inventory information from the point of con-

[1]The World Bank, USAID/Deliver Project, John Snow, Inc., Crown Agents and DFID

sumption using mobile devices; and (ii) mathematical optimization to determine the optimal quantity of drugs to ship to each health facility during each time period. This proposal is evaluated in §2.5 with a discrete-event simulation model that is validated with actual field observations obtained during the 2009 pilot experiment. Our results suggest that implementing our proposed alternative inventory control system would lead to substantial improvements in the availability of drugs at health facilities in Zambia. This work constitutes the primary motivation for a large-scale controlled field pilot of our proposed inventory system led by the MoH and supported by the World Bank, IBM and other partners which is planned to begin in three Zambian districts before the end of 2012. Further discussion of this upcoming pilot and concluding remarks are provided in §2.6.

ectionLiterature Review and Contributions

This paper studies an inventory policy for a system that supplies essential medicines to public sector health facilities in a resource poor environment. We review accordingly the literature in public health and operations management focusing on distribution of essential medicines in low income countries; impact of stock outs on health systems and health outcomes; and inventory distribution policies in multi-echelon systems.

Patel [1983], Yudkin [1978], Lall and Bibile [1978] highlight the problems with pharmaceutical access and affordability in low income countries and argue for a stronger public sector role in the provision of pharmaceuticals. Mamdani and Walker [1986] highlight problems in stock management, inventory record keeping, and supply ordering rendering the public distribution system for medicines ineffective in many developing countries. Quick [1982] highlights frequent shortages of medicines as a chronic problem impacting healthcare delivery in low income countries and study the impact of varying procurement frequency in public sector procurement. It shows how management science techniques can be best applied to improve availability of essential drugs. Sieter [2010] in his assessment of the design of better pharmaceutical policy for developing countries describes the results of a "dysfunctional" medicines supply chain. Many authors have highlighted the inefficiencies that result in the public sector medicine distribution system for HIV/AIDS medicines due poor information flows and poor system design [Windisch et al., 2011, Harries et al., 2007, Sued et al., 2011, Schouten et al., 2011]. Similarly multiple studies have been reported stockouts of malaria medicines in public health facilities [PLoS Medicine Editors, 2009, Zurovac et al., 2007, 2008, Kangwana et al., 2009].

Turning to the impact of supply chains, Kraiselburd and Yadav [2012] state that ineffective and poorly designed systems for purchasing and distributing medicines are one of the most important barriers to increasing access to medicines in low income countries. They attribute the ineffectiveness of the global health supply chain to coordination problems across multiple stakeholders with widely divergent objectives, poor supply chain design, and poorly designed myopic operational objectives. Roberts et al. [2004] present tools for diagnosing the performance of a pharmaceutical system and illustrate the role of structures, incentives and governance. Frost and Reich [2008] study the stakeholders and functions involved in making medicines, vaccines and other health technologies reach the populations they are intended for. WHO Maximizing Positive Synergies Collaborative Group [2009] highlighted the importance of supply systems in ensuring the overall success of new global health initiatives. They argue that while new global health initiatives have increased funding for the procurement of medicines, vaccines and other health technologies, the increase has not been matched by increase in the distribution of these supplies.

There are few studies to date that study supply chain redesign and measure its impact on availability or overall health system improvements. Sieter [2010] provides a review of various procurement and financing related interventions that help improve the pharmaceutical supply system in developing countries. Matowe et al. (2008) study the impact of training pharmacists in supply chain management on pharmaceutical availability in public sector drug supply chains. They do not find strong evidence that training achieves greater availability of drugs at the health facility level. Conn et al. (1996) describe the impact of a project to strengthen the basic management skills of district-level health teams in two out of the three health regions of the Gambia and find that it only leads to moderate improvements in health services delivery. Raja et al. (2008) use a simulation based optimization approach to redesign the public health distribution network in Kenya. Barrington et al. [2010] argue that lack of visibility of stock levels of malaria medicines at the health facility level is an important contributor to medicine stock-outs. They present results from a pilot SMS based system for providing visibility of health facility level stock of malaria medicines and show that it leads to fewer stock outs. Friedman et al. [2011] conducted a quasi-randomized experiment which compare two different supply chain structures against a control group and show that the supply chain with fewer tiers results in reduced stock outs at the health facility level. Bossert et al (2007) examine the impact of decentralization of specific functions within the medicines supply chain (e.g. needs quantification, inventory control, transportation, procurement and logistics management information systems (LMIS)) on multiple logistics system performance indicators. They show that decentralizing inventory control and LMIS to local authorities leads to poorer performance in these functions. In-depth assessment of the public medicines supply chain of six sub-Saharan African countries [WHO Maximizing Positive Synergies Collaborative Group, 2009, Yadav et al., 2011] shows that most of the countries record stock and consumption data at the district and health facility level on store ledgers, stock control cards and requisition forms. However, reporting such data to the higher levels of the distribution system for better supply planning is often difficult [Yadav et al., 2011]

Zipkin [2000], Porteus [2002] contain a review of the large stream of literature on inventory policy in multi-echelon systems, and Axsäter et al. [2002] contains a good discussion of the subset of these studies that focus on distribution systems. Within this literature, the problem we consider is distinguished by the assumption of lost (as opposed to backordered) unsatisfied demand, which seems required in a context where life-saving medicines are often distributed to patients who need to walk for several hours to reach care delivery facilities. Remarkably, the optimal replenishment policy is not even known for the single stage version of this problem with stationary demand and constant lead-times [Zipkin, 2008]. Our context however is further characterized by heterogeneous, non-stationary and stochastic demand and replenishment lead-times in a distribution system comprising a large number of facilities, and limited inventory available at the central warehouse that supplies them. We are not aware of any heuristic (let alone optimal policy) described in the literature for this problem. Among the stream of literature reporting heuristic policies adapted to realistic supply chain problems, Caro and Gallien [2010] is particularly relevant as it considers the determination of shipment quantities from a central store with limited stock to multiple retail locations in order to maximize overall sales. The present paper differs however from Caro et al. [2010] in the context and details of the distribution system considered, and consequently in the solution approach and approximations it uses.

In summary, the present paper makes the following contributions to the research literature on public health and operations management: (i) it identifies and quantifies the specific impact of a prevalent inventory control policy on patient-level drug stockouts in Zambia and several other countries in sub-Saharan Africa; (ii) it develops a detailed proposal for an alternative practical inventory control system and heuristic that are predicted to substantially improve patient access to drugs and other distribution performance metrics using a validated simulation model; and (iii) it provides a detailed description of a benchmark inventory problem and associated dataset arising in an important context that may motivate other researchers to conduct further related work.

2.2 Background on Zambia's Pharmaceutical Distribution System

2.2.1 Physical Flows

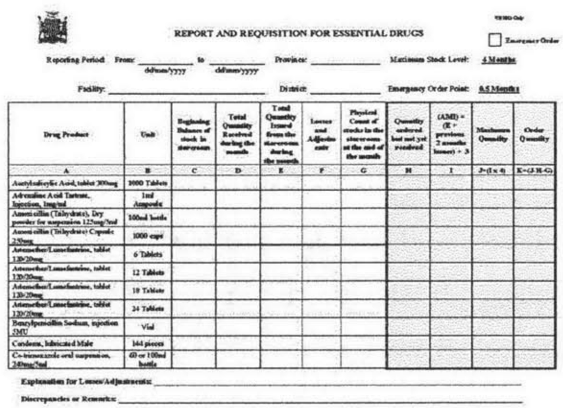
While essential medical drugs and supplies are sorely needed in Zambia, their mass distribution presents substantial difficulties due to low population density, poor road and communication infrastructure, and flood-related access cutoffs during the rainy season.

The three distribution channels for drugs in Zambia include the private sector, faith-based/mission organizations and the public distribution system [Yadav, 2007]. In part because of the limited spending power of Zambia's population (GDP per capita approximately $1000) however, the public distribution system has a predominant impact on public health. In that system, medicines are initially procured by the government with financial and technical assistance from external donors. Once received at the central warehouse in Lusaka, drugs and medical supplies are shipped to approximately 90 district stores and hospitals (*primary distribution*) by a para-statal agency called Medical Stores Limited (MSL). It relies on a monthly shipment cycle to satisfy replenishment orders which are determined by various stakeholders depending on the type of product. Upon receiving these orders, MSL carries out order picking and drugs are then shipped to the districts according to a pre-determined schedule of fixed truck routes.

Next drugs are distributed from the 72 district stores to Zambia's 1500 or so health centers, where both medicines and care are provided to patients for free. This *secondary distribution* is managed by the health ministry's District Health Offices (DHO) in partnership with health center workers. Depending on the district, some drugs may distributed by the DHO on a *cross-docking* basis, whereby it receives from MSL a package already prepared for its final health center destination that is just passed along for secondary distribution. Elsewhere and for other drugs, districts operate as intermediate stocking points that maintain their own dedicated inventory and prepare shipments for secondary distribution themselves. While the delivery of health center shipments may be performed in some districts according to a pre-determined schedule, most often secondary distribution rather works in an ad-hoc manner: in some instances the health center staff visits the district to pick up stock and in others the district team delivers the drugs to the health centers using its transportation capacity. Overall, secondary distribution is plagued by chronic shortages of both transportation resources and health center staff.

2.2.2 Information System and Inventory Control

The information and inventory control systems currently used in Zambia for the public distribution of medicines adhere to basic guidelines for logistics management designed and implemented by the USAID-funded DELIVER project [USAID | DELIVER, 2011a]. Zambia's logistics information system in most facilities is essentially paper-based. Its first key component is the MSL annual catalogue, which contains the list of all drugs and supplies available for order in various packaging formats from MSL. Its second key component is the monthly communication by fax, mail or hand delivery of replenishment order forms called *requests and requisitions forms*, or R&Rs for short (see Figure 2-1 for illustration).



REPORT AND REQUISITION FOR ESSENTIAL DRUGS

Emergency Order

Reporting Period: From: ________ to ________ Province: ________ Maximum Stock Level: 4 Months
dd/mm/yyyy dd/mm/yyyy

Facility: ________ District: ________ Emergency Order Point: 0.5 Months

Drug Product	Unit	Beginning Balance of stock in storeroom	Total Quantity Received during the month	Total Quantity Issued from the storeroom during the month	Losses and Adjustments	Physical Count of stocks in the storeroom at the end of the month	Quantity ordered but not yet received	(AMI) = (E + previous 2 months issues) ÷ 3	Maximum Quantity	Order Quantity
A	B	C	D	E	F	G	H	I	J=(I x 4)	K=(J-H-G)
Acetylsalicylic Acid, tablet 300mg	1000 Tablets									
Adrenaline Acid Tartrate, Injection, 1mg/ml	1ml Ampoule									
Amoxicillin (Trihydrate), Dry powder for suspension 125mg/5ml	100ml bottle									
Amoxicillin (Trihydrate) Capsule 250mg	1000 caps									
Artemether/Lumefantrine, tablet 120/20mg	6 Tablets									
Artemether/Lumefantrine, tablet 120/20mg	12 Tablets									
Artemether/Lumefantrine, tablet 120/20mg	18 Tablets									
Artemether/Lumefantrine, tablet 120/20mg	24 Tablets									
Benzylpenicillin Sodium, injection 5MU	Vial									
Condoms, lubricated Male	144 pieces									
Co-trimoxazole oral suspension, 240mg/5ml	60 or 100ml bottle									

Explanation for Losses/Adjustments: ________

Discrepancies or Remarks: ________

Figure 2-1: Request and Requisition Form (R&R), 2010 Version.

The communication of these forms goes in the opposite direction to the flow of goods: in districts operating under cross-docking, each health center submits its monthly R&R form directly to MSL. Elsewhere, each health center and hospital thus sends a R&R to its district pharmacy (DHO) monthly, and each DHO sends in turn a R&R monthly to MSL. In order to coordinate the transmission of this information, each echelon faces a monthly deadline for transmitting its R&R upstream, which is based on the set schedule of deliveries from MSL to the districts.

As seen in Figure 2-1, these order forms prompt users to provide a number of information variables (indexed from A to K and listed in columns) for each one of the drugs carried by the corresponding facility (listed in rows). These variables fill two purposes: (i) they document the aggregate impact of inventory transactions (receipt, deliveries, counting adjustments) on local stock (i.e., the various components of the stock balance equation); and (ii) they support the calculation of the replenishment order from the facility filling the R&R form to the one receiving it, which appears in the last column of the form (variable K). That is, the R&R is an order form designed to provide partial downstream inventory visibility, and help enforce local compliance to the following ordering rule (so-called *min/max* rule) whereby regular monthly order quantities are given by

$$\text{Order Qty} = (\ M \text{ x Avge Monthly Issues}\ ) - \text{Stock on Hand} - \text{Undelivered Past Order Qty}, \tag{2.1}$$

where:

- M is the so-called *maximum stock level*, representing the number of months of recent past consumption to which inventory should be replenished. While the value of M is shown as 4 months in Figure 2-1, it more generally ranges from 2 to 5 months depending on the type of facility;
- **Avge Monthly Issues** (noted AMI in column I of Figure 2-1) is the average of monthly quantity issued for that drug at this location, calculated over the previous three months. As a result, the first term M x **Avge Monthly Issues** (column J in Figure 2-1) in the r.h.s. of (2.1) effectively corresponds to the base stock level used by that policy;
- **Stock on Hand** (column G in Figure 2-1) is the physical count of stock of that drug that is physically present in the facility's storage room at the time when the order is calculated. Note that this does not include any stock present in the care delivery area of that facility, if applicable;

- **Undelivered Past Order Qty** (column H in Figure 2-1), when applicable, is the total quantity of that drug previously ordered by that facility but not yet delivered. Note however that some districts do not include this quantity in their calculations, depending on the version of the R&R form that they use.

In summary, inventory control in Zambia relies on a monthly review base stock policy supported by a paper-based information system. Finally, a first-come-first-serve (FCFS) allocation rule is most often followed in situations where not enough inventory is available to satisfy all replenishment orders received at a given location (either a district or MSL). That is, individual shipment requests are filled in the sequence they arrive until no more inventory is available, at which point the current replenishment order being considered may be filled only partially and the subsequent ones not at all.

2.2.3 Public Sector Pilot

In order to reduce the stockouts of essential drugs in its public health facilities (see §2.1), the government of Zambia together with the World Bank, USAID and other partners conducted between April 2009 and April 2010 a landmark randomized pilot experiment comparing two possible alternative distribution models against the current one [Friedman et al., 2011]. Specifically, this experiment was designed to evaluate the impact of two different supply chain structures (intermediate stocking and cross docking) complemented by a new employee position (*commodity planner*) dedicated to inventory management at the district level. These two distribution models were each implemented in 8 randomly selected rural and semi-rural districts, and an additional eight districts were selected for observation as a control group. As discussed in Friedman et al. [2011], the first of these interventions (cross-docking) resulted in a substantial and statistically significant improvement of the availability of drugs to patients during the fourth quarter of 2009 (the evaluation period considered in that study) as compared to the same period in the previous year and the control districts. On the basis of these results, the government of Zambia decided in late 2010 to initiate a progressive deployment of cross-docking to the rest of the country.

Beyond the supply chain modifications considered as part of these interventions, this pilot experiment also provides an opportunity to better understand the specific performance of the inventory control method (2.1) used throughout. This is because the commodity planners employed in the intervention districts ensured a high level of adherence to that ordering rule. The following section describes our works towards that goal.

2.3 Performance of Existing Inventory Control System

We restrict our performance analysis to the cross-docking intervention, which was shown to be superior by the pilot experiment and is currently being deployed more widely in Zambia. We also focus on the anti-malarial medicine Artemether/Lumefantrine (brand name Coartem). Firstly, Artemether/Lumefantrine (AL) is important to global public health as the recommended first-line treatment for malaria in many countries including Zambia[2]. Second, the demand for AL is seasonal because malaria incidence is highly correlated with rainfall patterns and Zambia experiences a marked rainy season between December and March. Third, AL is distributed to all care delivery facilities in Zambia. AL is thus an important product in itself but is also representative of the health commodities that are most challenging from an inventory control standpoint.

2.3.1 Methods

To more accurately evaluate the performance of the existing inventory control system, we supplemented the survey data pertaining to Q4 2009 that was collected during the

[2]There were approximately 225 million cases of malaria and 781,000 related deaths worldwide in 2009. The burden of malaria falls primarily on countries in sub-Saharan Africa, with the majority of deaths being young children. Malaria is also a major hindrance to the economic development of these countries [World Health Organization, 2011].

pilot with data from the stock control cards (henceforth *stock cards*) located in the drug storage area of each health center. Stock cards provide a complete record for each drug in the pharmacy of the inventory level, receipts from the DHO, and issuances to the HC dispensary over time (see Figure 2-2).

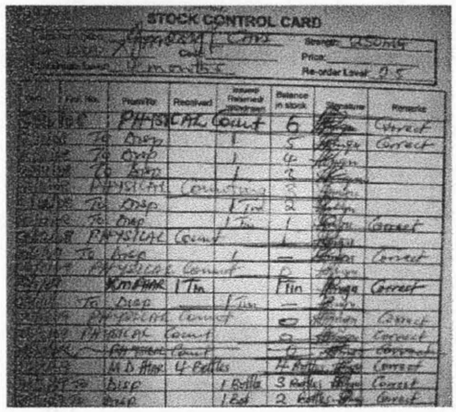


Figure 2-2: Example of a Stock Control Card

Note. A stock control card for the drug amoxycillin, taken from the Jimbe health facility in the Mwinilunga district.

AL comes in four different pack sizes (6, 12, 18 and 24), with the numbers indicating the quantity of pills included in a tablet constituting a single dose for a patient. The dosage for a patient is dependent on his/her body weight, with the pack size of 24 for adults and smaller ones for children with smaller body weights. Any box of AL contains 30 individual doses regardless of pack size.

1649 digital photographs of their stock cards in 121 clinics (25 in intermediate stocking districts, 96 in cross-docking districts) were taken by commodity planners during their regular district tours between May 2009 and June 2010. Database transcription was performed by specialized data entry firm DDD using a double-key process. Consistency of database with inventory balance equation was verified.

A subset of 12 clinics located in 4 cross-docking districts with complete time coverage of one year or more in the dataset was selected for demand estimation purposes. In each of these clinics, raw demand rate on any day with at least one box in recorded inventory was computed for every product as the quantity of tablets in a box divided by the number of days between the last and next issues of a box. To smooth the resulting discontinuities and estimate censored demand, a triple centered moving average operator with successive half-widths 40, 30 and 20 days of non-censored data was applied [Makridakis et al., 1998]. Final total demand estimate for AL in each clinic was obtained as the sum of the smoothed uncensored demand estimates for individual AL products.

2.3.2 Results and Discussion

Figure 2-6 provides an example of the detailed historical data of issues, demand and inventory that we were able to obtain for 12 HCs located in cross-docking districts throughout Zambia. Because these data series were obtained by aggregating all four AL pack sizes, any period without inventory seen in the graph corresponds to an actual stockout at the HC pharmacy of all AL pack sizes simultaneously. Figure 2-3 shows stockout and inventory data aggregated over these 12 HCs. We observe the following: as independently found as part of the pilot evaluation, cross-docking HCs experience very few stockouts during the fourth quarter of 2009. This is however not representative of the service level during other times of the year, and in particular a substantially higher stockout rate is observed between mid-February to mid-April 2010. Consistently, the median HC inventory level drops to almost zero during this latter period, even though it reaches 6 months of future demand at the beginning of the pilot in July 2009 (most likely due to special shipments made as part of the pilot set-up activities).

Further highlighting the effect of non-stationary demand, Figure 2-4 shows a rescaled version of the data seen in Figure 2-3 where the origin of the time series corresponding to every individual health center has been set independently to the estimated demand peak date for that location before aggregation statistics are computed. This re-scaling effectively removes the effect of demand pattern heterogeneity across health centers when interpreting stock-out dynamics. We observe the following: (i) While service levels are quite good in the months before peak demand, a high rate (15–20%)

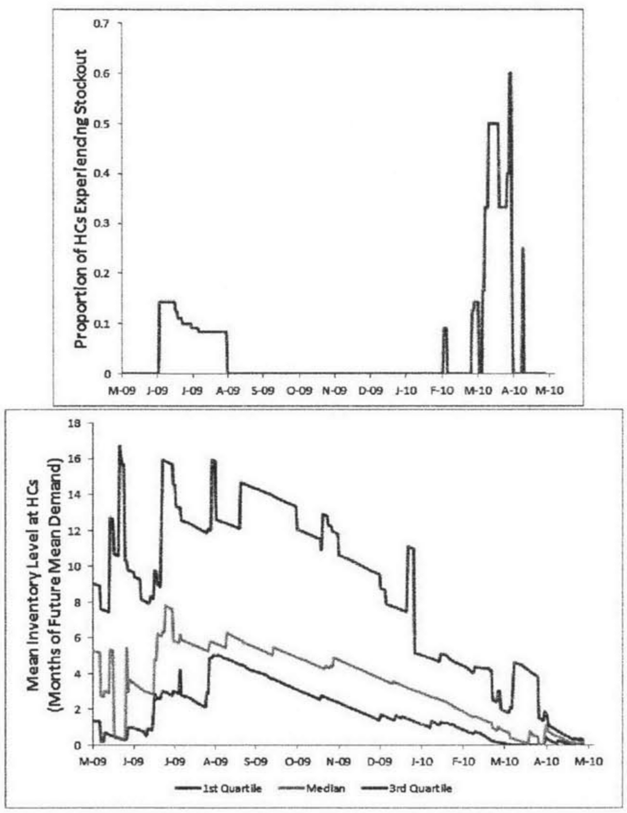


Figure 2-3: Health Facility Stock-Outs and Inventory Levels

Note. Stockouts and inventory levels, aggregated over twelve health facilities from cross-docking districts.

of health centers stock out in the following three months; and (ii) inventory levels are fairly stable during the period of two to six months before peak demand, but quickly fall during the two months preceding peak demand and remain quite low during the three following months.

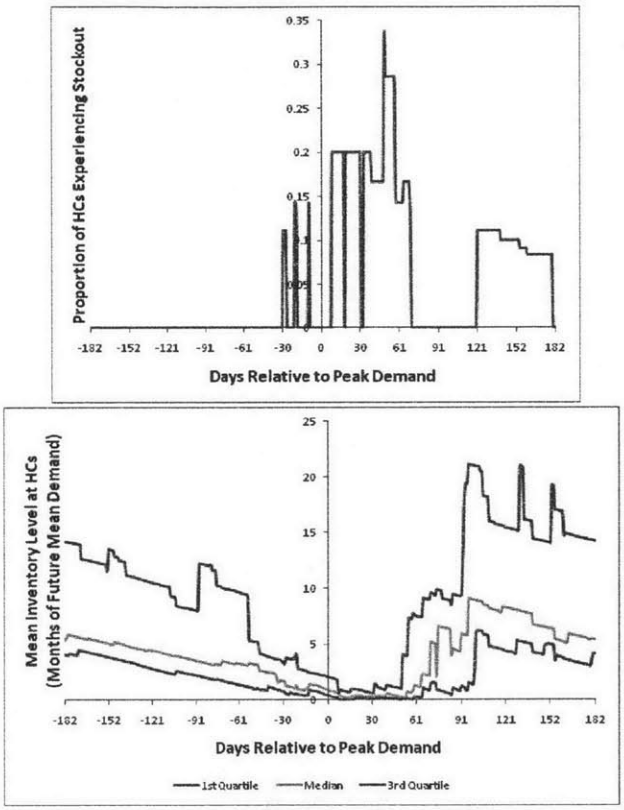


Figure 2-4: Health Facility Stock-Outs and Inventory Levels (Rescaled)

Note. Stockouts and inventory levels, aggregated over twelve health facilities from cross-docking districts and re-scaled in time. The time re-scaling is based on the assumption that demand at a health facility has yearly seasonality; therefore we can shift the time axis at each health facility cyclically, so that day zero is the day of estimated peak demand. Following this, we aggregated the results over the HCs to obtain these graphs.

A striking aspect of these drug stockouts is that they occured even though (i) the concurrent employment commodity planners and central monitoring staff guaranteed a high level of adherence to the recommended inventory control policy (2.1); and (ii) as MSL inventory records show, there was ample stock of that drug in the central warehouse during the entire relevant time period (Q4 2009, Q1 and Q2 2010).

These observations are fully explained by a relatively straightforward analysis of equation (2.1): the replenishment target $M\times$ Avge Monthly Issues, which determines shipments meant to cover future demand, is based on the average *past* consumption rate over the previous three months. As a result, these shipments systematically fail to anticipate predictable upcoming demand changes in either direction, as well as any upcoming changes in delivery lead-times. When demand starts to substantially increase in the couple of months preceding a demand peak, the quantities ordered reflect previous lower consumption rates and are therefore insufficient to cover the upcoming demand surge. The safety stock resulting from the multiple M takes some time to deplete however, so that the first stockouts only appear in the second half of the peak demand period, as seen in Figure 2-4. As further discussed in §2.5.1, we were able to reproduce these same periodic stockout dynamics at the HC level using a discrete-event simulation model capturing the replenishment policy (2.1) under the assumption of unlimited inventory availability at MSL. This lends even greater support to the interpretation provided above and to the specific impact on stockouts of the basic min/max inventory control policy (2.1) that we highlight and quantity in this paper.

2.4 Alternative Inventory Distribution System

2.4.1 Qualitative Considerations

As discussed in §2.3.2, a key reason for the relatively poor performance of the current system is the use of average monthly consumption over the past three months as a predictor of the demand over the next replenishment period. This implicitly amounts to the untenable assumptions that there are no stock-outs and that both demand and lead-times are stationary. Therefore, a key step towards better inventory control is to generate forecasts for patient demand and supply lead-times reflecting the substantial seasonality of these quantities. Because of staffing and infrastructure limitations affecting public health workers in Zambia however, it does not seem practical in the short term for these forecasts to be generated at the health center nor even the district level. That observation dictates in turn a vendor-managed inventory (VMI) design whereby the primary contribution of local health workers to inventory control is to facilitate continuous and accurate visibility of their stock levels by the national warehouse (and perhaps transmit ad-hoc local information about future demand and accessibility conditions), while shipments to each facilities are determined at the central level using forecasts constructed from that information.

In such system, the central problem of frequently computing appropriate shipment quantities for a large network, when overall supply is occasionally limited and both demand and supply lead-times are seasonal and heterogeneous across facilities, is quite complex. That problem is also sensitive, because in situations of scarce inventory its solution may have an impact on public health, and that impact may differ across districts. Both considerations suggest the use of an optimization model for calculating shipments, because that methodology has a proven track record for solving similarly complex industrial distribution problems (Caro et al. [2010], Foreman et al. [2010]), and because it provides an objective justification for the generated solutions. As discussed in §2.1 however, we are not aware of any heuristic described in the literature for the specific inventory distribution problem with non-stationary demand and lead-times that is relevant in this context, and thus describe next such a heuristic in §2.4.2 that we then proceed to evaluate through simulation in §2.5.

Finally, the amount of information required to centrally determine appropriate shipment quantities (quantities issued and updated stock levels at the delivery points for many drugs, dates of shipment receipts) and the staffing challenges encountered at the health centers justify the consideration of a digital data transmission system between the clinics and the central warehouse that would minimize labor requirements associated with data entry and communication. Given the current limitations of power and communications infrastructures in rural Zambia however, the only viable option for such system in the foreseeable future appears to be a smart phone with a client application for ergonomic data entry (possibly supplemented by a bar code scanner) and access to the cellular phone network for transmission purposes. While SMS-based solutions have been tested in similar environments for a limited number of drugs [Barrington et al., 2010], because of ergonomic considerations they seem inappropriate for handling the range of essential drugs carried in Zambia's public health centers.

2.4.2 Shipment Optimization Heuristic

We focus here on a possible optimization model formulation for the central shipment calculation problem discussed in §2.4.1, and provide later in §2.5 a numerical study of its likely performance in Zambia when used as part of the proposed inventory distribution system just presented. That optimization model is an inventory planning linear program (LP) that is independently instantiated and solved on a rolling horizon basis for every drug to be distributed, similar in spirit to the model described in Foreman et al. [2010]. Its primary decision variables are the quantities of the drug considered to be sent to each health center as part of each monthly shipment scheduled over the planning horizon; its objective is to minimize a weighted combination of expected lost demand and inventory holding costs calculated over the entire planning horizon and geographic region considered. A planning horizon of six months to a year is appropriate in light of the seasonality patterns of demand and delivery lead-times in this environment, and a planning period of one week appears suited to the monthly delivery cycle used by MSL and likely input data accuracy. While future periods in the horizon are considered to prevent any myopic behavior, upon any model run only shipment variables corresponding to the next scheduled delivery are meant to be implemented. Finally, this model considers the deliveries to the central warehouse as exogenous input data, because the upstream procurement process is handled by a different organization, the relevant considerations (contractual agreements, production capacity constraints, multiple suppliers) are quite different, and the corresponding necessary data is harder to acquire in practice. An exact model definition follows.

Sets and indices:

- $\mathcal{H}$ set of final shipment destinations (health centers) considered. The central stocking point (MSL warehouse) is denoted M.
- $\mathcal{T} = \{T_0, ..., T_1\}$ set of consecutive discrete periods (weeks) in the planning horizon, where T_0 and T_1 are the first and last periods in the horizon, respectively;
- $\mathcal{K}_t^h$ set of approximating tangents for the lost demand function of health center $h \in \mathcal{H}$ in period $t \in \mathcal{T}$.

Input data:

- $D_{T_0,t}^h$ expected demand at health center h during week t estimated in week T_0. Based on a maximum likelihood estimation of demand distributions from the stock card data discussed in §2.3.1, it is assumed that the distributional forecast $\mathbf{D}_{T_0,t}^h$ of demand in week t available in week T_0 is lognormal with mean $\bar{D}_t^h$ and generated according to the multiplicative martingale model of forecast evolution (MMFE). More details on demand estimation and forecast generation are provided in §2.5.1 and §2.5.1, respectively.
- $L_t^h[\beta]$ β-conservative deterministic lead time from MSL to health center h for a shipment initiated in week t, where β is a parameter in $(0, 1)$. The β-conservative lead time sets the first shipment at or after t from MSL to health center h and any shipments currently in the pipeline to the minimum possible lead time, while setting any subsequent shipments from MSL to health center h equal to the β-fractile of the lead time distribution. The intuition behind using β-conservative lead times is the following: By making the first shipment and pipeline shipments arrive as soon as possible, and the subsequent shipments arrive late, the first shipment quantity has to tide the health center until the second shipment arrives, so this tends to make the first shipment larger than using deterministic lead times equal to the mean.

- $I_{T_0}^h, I_{T_0}^M$ initial inventory levels at health center h and MSL, respectively;
- $X_{T_0-2}^h$ current pipeline shipment quantity from MSL to health center h which has already been received by the district;
- $X_{T_0-1}^h$ current pipeline shipment quantity from MSL to health center h which has not been received by the district yet;
- X_t^M quantity planned to be delivered at MSL by suppliers in week t;
- A_{tk}^h, B_{tk}^h slope and intercept of approximating tangent $k \in \mathcal{K}_t^h$ for the lost demand function of health center h during week t estimated as of week T_0. Since that lost demand $\mathbf{E}[(\mathbf{D}_{T_0,t}^h - i_t^h)^+]$ is a convex function of the starting inventory level i_t^h (see variable definition below), it can be approximated arbitrarily closely by the upper enveloppe of a discrete set $\mathcal{K}_t^h$ of its tangents. Their slopes and intercept are calculated using the closed form expressions for the lost demand function and its derivative that are available under the distributional assumptions for $\mathbf{D}_{T_0,t}^h$ (see §2.5.1);
- C penalty cost for one unit of lost demand relative to the cost of holding one unit of inventory for one period at a health center.

Decision variables:

- x_t^h quantity to be shipped from MSL to health center h during week t. These variables are only defined where appropriate, i.e. according to the pre-determined schedule of truck routes for primary distribution whereby each district only has a single monthly shipment opportunity (see §2.2.1);
- i_t^h, i_t^M approximate expected inventory level at the beginning of week t in health center h and MSL, respectively;
- s_t^h approximate expected shortages (lost demand) at health center h during week t

Formulation:

$$
\begin{aligned}
\min \quad & \sum_{h\in\mathcal{H}}\sum_{t=T_0}^{T_1}(C\times s_t^h + i_t^h) && (2.2)\\
\text{subject to} \ : \quad & i_{T_0}^M = I_{T_0}^M && (2.3)\\
& i_{t+1}^M = i_t^M + X_t^M - \sum_{h\in\mathcal{H}} x_t^h \quad \forall t\in\mathcal{T} && (2.4)\\
& i_{T_0}^h = I_{T_0}^h \quad \forall h\in\mathcal{H} && (2.5)\\
& i_{t+1}^h = i_t^h - D_{T_0,t}^h + s_t^h + \sum_{u\in\{T_0-2,\ldots,t:u+L_u^h[\beta]=t\}} x_u^h \quad \forall h\in\mathcal{H}, t\in\mathcal{T}\backslash\{T_1\} && (2.6)\\
& s_t^h \le D_t^h \quad \forall h\in\mathcal{H}, t\in\mathcal{T} && (2.7)\\
& s_t^h \ge A_{tk}^h i_t^h + B_{tk}^h \quad \forall h\in\mathcal{H}, t\in\mathcal{T}, k\in\mathcal{K}_t^h && (2.8)\\
& x_t^h = X_t^h \quad \forall h\in\mathcal{H}, t\in\{T_0-2, T_0-1\} && (2.9)\\
& x_t^h, i_t^M, i_t^h, s_t^h \ge 0 \quad \forall h\in\mathcal{H}, t\in\mathcal{T} && (2.10)
\end{aligned}
$$

The objective (2.2) captures the sum of lost demand and inventory holding costs over the planning horizon and the set of health centers considered; constraints (2.3)-(2.6) and (2.5)-(2.6) are the inventory balance equations for MSL and every individual health center, respectively; constraints (2.7)-(2.8) implement the linear piecewise approximation of the lost demand function; constraint (2.9) ensures that shipment decisions that have been made in the past are correctly taken into account; and constraint (2.10) ensures that all decision variables are non-negative, which together which (2.4) implies that total shipments in any period do not exceed the inventory available at MSL then.

2.5 Performance Evaluation

Our performance evaluation study is designed to answer the following questions: (i) What is the performance of the proposed inventory distribution policy described in §2.4, both in absolute terms and relative to the inventory control policies currently used in Zambia (see §2.2). In particular, what would be the likely impact observed if our proposed policy were implemented in that country? (ii) What is the explanation for any potential performance differential observed? In particular, what is the specific performance impact of environment features such as inventory scarcity, forecast quality and communication delays, and policy parameters such as the maximum stocking level(s) (cross-docking and intermediate stocking flows relying on min/max policies) and lost demand penalty / lead time fractile (for the proposed policy)?

The specific metrics we use to evaluate the performance of all distribution policies considered over any time period are the following:

Mean service level: The proportion of patient demand satisfied from available inventory at all the HCs in the network considered.

Mean inventory level: The average on-hand inventory level at the HCs in the network considered, expressed in weeks of average demand. This is relevant to assess the trade-off achieved between service level and inventory (for drugs with limited shelf lives holding costs possibly include waste due to expiration), but also because storage at the health centers may be poor or limited.

Standard deviation of service level: The standard deviation of service levels calculated independently for each HC across all sites in the network considered, reflecting distribution fairness or the degree to which access to drugs may differ for patients depending on their home location.

2.5.1 Methodology

Simulation Model Structure and Scope

We use a discrete-event simulation model with a weekly time period predicting on-hand inventory dynamics of all combined AL products in a network comprising the central warehouse, 12 DHOs and 212 HCs. This geographic coverage amounts to approximately 17% of Zambia's facilities and corresponds to the districts for which demand and lead time data could be collected or estimated by leveraging the presence of a commodity planner during the public sector pilot (see §2.2.3). We note that these districts were carefully selected as a subset that is representative of the entirety of Zambia as part of the experimental design of this pilot [Friedman et al., 2011]. The sequence of events simulated by this model in each period is the following:

1. Planned receipts for this period are credited to the on-hand inventory of each location;

2. Demand in each health center is generated according to the stochastic demand model described in §2.5.1 and debited from the local on-hand inventory. Lost demand is recorded in case demand exceeds available inventory;

3. Shipments from the central warehouse to the set of facilities on the shipment schedule that week are computed according to the inventory policy being simulated (see §2.5.1) and debited from the central warehouse inventory (or district office inventory if appropriate). Reflecting the actual pre-determined monthly primary distribution schedule of fixed truck routes covering each a subset of districts (see §2.2.1), the simulated districts are evenly partitioned into subsets associated with single monthly shipment opportunities that are regularly spread through the month. Lead times for these shipments are generated according to the stochastic lead time model described in §2.5.1 and the corresponding planned receipts are added to the list of future events;

In particular, this simulation model captures key features of Zambia's distribution system including the predictable and unpredictable variability associated with both demand and shipment lead times, the monthly order and shipment schedule at MSL, the scarcity of central inventory and potential communication delays from the DHOs and HCs. Among key assumptions, it considers the receipts of inventory by MSL as exogenous input data (see §2.4.2 for discussion) and assumes that demand and inventory for the various pack sizes of AL are fully substitutable, consistent with our personal communications with warehouse managers and field observations[3]. Crucially, we were able to validate the predictive accuracy of this model by comparing its output with empirical stock out measurements observed under known operating conditions, as further discussed in §2.5.1.

Demand Model

A key input of our simulation model is the probability distribution of demand for all AL products at each one of 212 HCs every week of the year. We constructed this demand model in two steps.

First, we constructed estimates of expected weekly demand through an entire year at 18 HCs for which we had been able to collect sufficient stock card data through digital photographs[4], using the methodology described in §2.3.1. By normalizing these curves so that the mean weekly demand equals one, we thus obtained 18 different seasonality patterns associated with these specific locations spread throughout Zambia (see Figure 2-5).

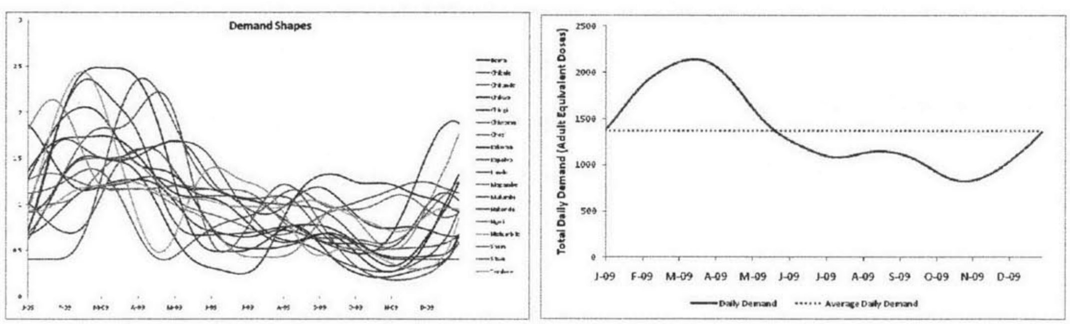


Figure 2-5: Estimated Demand Seasonality for AL

Note. (Left) Demand seasonality at 18 HCs, with the mean daily demand at each HC scaled to 1. (Right) Estimated seasonality of total demand (summed over all HCs in the twelve districts).

We also fitted various parametric distributions to the differences between the mean of

[3]The different pack sizes of AL are characterized by different numbers of pills on each tablet (which constitutes a dose), however individuals pills from different pack sizes are identical. Two tablets of 6 pills each are therefore rigorously equivalent to one tablet of 12, etc.

[4]These 18 facilities include the 12 HCs from the cross-docking districts for which data is presented in §2.3.2, and another 6 HCs from the intermediate stocking districts (see §2.2.3).

weekly demand thus estimated and the observation of actual consumption during weeks with no stockouts. Using maximum likelihood, we found that the negative binomial and lognormal distributions with a constant coefficient of variation of 0.5 fitted our dataset best (because of its theoretical connection with the multiplicative martingale model of forecast evolution, we selected the lognormal). Because the distances between pairs of locations selected from this subset spanned a wide range of values, we could also perform a statistical test to estimate the distance threshold below which any two pairs of locations were more likely to have identically distributed peak demand weeks than not (about 100km in this dataset). Finally, we also performed a linear regression of the total yearly demand for AL (expressed in equivalent adult doses) thus calculated at each of these HCs against the independent variables of *catchment population* and local *malaria incidence* obtained from an existing 2006 epidemiological survey of Zambia's health centers obtained from the Ministry of Health, and *average patient visits per day* which was collected as part of a survey of local health personnel conducted by all commodity planners in their respective districts in the Fall of 2009. Despite the seemingly coarse and subjective nature of the last independent variable, this regression model yielded a relatively high fit ($R^2 = 0.74$) over the restricted set of 18 locations.

	Estimate	Standard Error	t-value	p-value
Intercept	−2.745e3	7.91e2	−3.444	0.003955
Catchment	1.951e−1	4.053e−2	4.813	0.0000276
Patients × Incidence	9.625e−2	3.513e−2	2.740	0.015944
Incidence	2.539	1.146	2.215	0.043816

Table 2.1: Estimation Metrics for Annual AL Demand

As a second step, we leveraged the various estimates obtained from the detailed study of these 18 HCs in order to construct weekly demand distributions for the remaining 194 HCs covered by our simulation model: (i) we assigned to each HC one of the 18 estimated demand seasonality patterns probabilistically, using assignment probabilities decreasing with distance and reflecting the 100km similarity threshold mentioned above; (ii) we computed the prediction of total yearly demand for AL products provided by the regression model just mentioned by substituting the specific values of the independent variables for that HC; (iii) we multiplied each one of the normalized seasonality factor obtained in (i) by the weekly demand average resulting from (ii); (iv) we assumed a log-normal distribution of weekly demand with a mean given by (iii) and a coefficient of variation equal to 0.5.

Lead Time Model

A second key input of our simulation is the probability distribution of shipment lead times from the central warehouse to each DHO (primary distribution) and from each DHO to each HC (secondary distribution). Because primary distribution relies on an appropriately sized fleet of modern and well-maintained trucks traveling on tar roads and many personal communications suggested it was quite reliable, we assumed for that part a deterministic lead-time equal to two weeks (spanning shipment computation, picking, packing and loading operations at MSL, transportation, and receiving operations at the DHO).

For several reasons including insufficient transportation capacity in the districts and occasional lack of planning, secondary distribution lead times were reported to be much longer and variable. In addition, lead times to some specific HCs were strongly affected by seasonal accessibility problems due to road cut offs during part or all of the rainy season. Our approach was to first construct DHO-specific stationary lead time distributions for the times of the year and HCs without accessibility problems, and then convolute these base distributions when and where appropriate with a second HC-specific and non-stationary distribution specifically capturing the impact of local seasonal accessibility problems.

The two pieces of data available to us for estimation purposes were (i) reports submitted monthly by all commodity planners to MSL between June 2009 and June 2010 including the dates when HC orders were received by the DHO and the DHO made a delivery to each HC; and (ii) a survey of all HC workers conducted by commodity planners in their respective districts and included their estimates of the probability that the road from the DHO to a HC would be inaccessible in any week of a given month.

We used (i) to construct for each district a dataset of lead times for all HCs and time periods which were not affected by accessibility issues according to (ii). Each DHO-specific stationary distribution was then obtained as the empirical distribution of lead times over this dataset. To construct the HC-specific non-stationary distribution capturing accessibility-related delays, we interpreted the probability estimates a_t^h from (ii) associated with each HC h in week t as the probabilities of successive independent Bernoulli draws that any delivery normally attempted that week would be delayed by at least one more week. That is, whenever the realization of the DHO-specific stationary lead time suggests a delivery received at HC h in week t, then the conditional probability that the shipment will actually be received in week $\tau \geq t$ given local seasonal accessibility problems is given by

$$(1 - a_\tau^h) \prod_{k=t}^{\tau-1} a_k^h,$$

where the product operator over an empty/undefined set is equal to 1.

Policies and Scenarios

The three families of inventory distribution policies we evaluate are denoted IM (intermediate stocking at the district with min/max inventory control at the DHOs and HCs), XM (cross-docking at the district with min/max inventory control at the HCs) and XO (cross-docking at the district with proposed optimization-based inventory control). Within each family, a given policy is characterized by family-specific inventory control parameters such as the maximum stock level (min/max policies) or the lost demand penalty / lead time fractile (proposed policy), and the following two infrastructure variables:

delay : This variable represents the communication time delay expressed in weeks for any information (order and/or inventory position) from a downstream level of the supply chain to be received by the next upstream level. In particular, the current system involving paper-based information transmission from the HCs to the DHOs and the DHOs to MSL is represented reasonably well by the value $delay = 1$. A system involving the ongoing transmission of inventory receipts and issues by HCs using mobile phones as described in §2.4.1 would in contrast be characterized by $delay = 0$.

$forecast$: This variable represents the method available to forecast demand at any location when computing inventory replenishment quantities. The method implicitly underlying the basic min/max policies currently used, which consists of forecasting the average monthly demand over the next few months with the average of monthly issues over the previous three (see §2.2.2), will be referred to with $forecast = ami$. In order to capture more common and advanced forecasting methods in environments with seasonality, we use the standard multiplicative form of the martingale model of forecast evolution (MMFE) described in Heath and Jackson [1994], which seems adapted to our environment where demand can be modeled reasonably well by lognormal distributions with a constant coefficient of variation (see §2.5.1). Specifically, distributional forecasts available in period s for the demand in period $t \geq s$ are generated by the process

$$\mathbf{D}_{s,t} = \bar{D}_t \exp(\epsilon_{t,t} + \epsilon_{t-1,t} + \dots + \epsilon_{s,t}) \exp(\boldsymbol{\epsilon}_{s-1,t} + \dots + \boldsymbol{\epsilon}_{t-H,t}), \tag{2.11}$$

where $\bar{D}_t$ is the demand mean obtained from the estimation procedure described in §2.5.1, H is the length of the forecast horizon, and $\epsilon_{u,t}$ are normal random variables representing the uncertainty that is revealed in period $t - u$ concerning demand during period t (bold characters are used to differentiate yet unknown random quantities from their known realizations)[5]. Consistent with (2.11), the simulated

[5]If $\epsilon_{u,t} \sim N(\mu_u, \sigma_u^2)$, then parameters μ_u and σ_u^2 are constrained by $\mu_u = -\sigma_u^2/2$ so that $\mathbb{E}[\exp(\epsilon_{u,t})] = 1$. Others constraints stem from the specified moments of demand $D_{t,t}$ and sched-

realization of demand in period t is generated as $\bar{D}_t exp(\sum_{u=t-H+1}^{t} \epsilon_{u,t})$, which is indeed lognormal. We consider two cases of forecast quality within this model, which arguably constitute optimistic and pessimistic scenarios for the predictive accuracy that could be achieved in Zambia by implementing a forecasting software relying on standard time-series analysis. The scenario $forecast = myo(pic)$ corresponds to the case $H = 0$ where forecasts do not improve over time because no uncertainty about demand is revealed in the preceding periods; The case $forecast = ind(ustry)$ essentially corresponds to the best performing "statistical method" discussed in Heath and Jackson (1994) and for which necessary parameters are provided including $H = 3$ months and 44%, 30%, 18% and 7% of demand variability resolved 3, 2, 1 months before sales and during the month of sales, respectively.

Fully specified policies can thus be referred to using the following notation:

- $IM_{M_{DHO}}^{M_{HC}}[delay, forecast]$: The intermediate stocking policy with min/max inventory control and maximum stock levels set to M_{HC} months at the HCs and M_{DHO} months at the DHOs, and infrastructure variables *delay* and *forecast* as discussed above. Consistent with observed practice, whenever the inventory available to any supplier (MSL or DHO) for fulfilling all orders in a given week is insufficient, it is allocated based on a fixed priority ordering of DHOs or HCs. In addition, the replenishment schedule of DHOs corresponds to the pre-determined monthly schedule of truck routes from MSL (see §2.2.1) and replenishment orders for the HCs are all submitted in the first week of each month.

- $XM^{M_{HC}}[delay, forecast]$: Definition of the cross-docking policy with min/max inventory control analogous to that of $IM_{M_{DHO}}^{M_{HC}}[delay, forecast]$, except that the specification of the maximum stock level M_{HC} is only required for the HCs. In addition, the replenishment schedule of HCs is synchronized with the pre-determined monthly schedule of truck routes from MSL.

ule of demand uncertainty resolution with u.

- $XO^C_\beta[delay, forecast]$: The cross-docking policy with shipment quantities calculated through the proposed inventory distribution heuristic stated in §2.4.2, with a lost demand penalty equal to C and lead time fractile parameter equal to β. The planning horizon T_1 for that policy is set to 36 weeks.

For example, $IM^2_3[1, ami]$ and $XM^4[1, ami]$ represent the min/max inventory distribution policies utilized in the intermediate stocking and cross-docking interventions of the public sector pilot, respectively (see §2.2.3), and $XO^{10}_{0.8}[0, ind]$ would represent the version of our proposed inventory distribution policy with a communication system based on mobile phones, a forecast accuracy competitive with documented industry benchmarks, a deterministic approximation of shipment lead-times at fractile 0.8 and a lost demand penalty parameter equal to 10.

We note that for the inventory distribution problem characterized by the constraints captured by the simulation model and the objective function defined by (2.2), a lower bound is achieved by the ***clairvoyant*** shipment policy obtained by solving problem (2.2)-(2.6) and (2.9)-(2.10) with a planning horizon equal to the simulated one and the values of demand and lead times fractiles replaced with their actual simulated realizations. That is, the (unrealistic) policy obtained when assuming that all future demand and lead times are perfectly known upfront. In the following we will denote this policy by CL^C and use it as a performance benchmark to bound the suboptimality of the other non-clairvoyant policies we evaluate.

Finally, the key environment scenario variable that we consider is the scarcity of inventory available at MSL relative to network-wide demand at the HCs. Specifically, we assume one delivery of Q^{MSL} units of inventory at MSL every 3 months, and vary Q^{MSL} to achieve different values of the ***suppy/demand ratio*** defined as

$$\frac{4Q^{MSL}}{\sum_{t=1}^{52}\sum_{h\in\mathcal{H}}\bar{D}^h_t},$$

where the denominator represents the sum over all HCs of average simulated demand through one year.

Validation

We first performed a qualitative validation of our simulation model by comparing the sample paths of inventory and lost sales it generated with the actual historical traces for these same quantities that we had been able to estimate directly from the stock cards for a subset of HCs (see §2.3.2). As illustrated by Figure 2-6 for a representative HC, the simulation model did indeed reproduce qualitatively the same behavior over time. Namely, that figure shows that even in a situation with unlimited inventory at MSL, simulated stockouts also tend to appear in the second half of a peak demand period (from January to April in Figure 2-6). In addition, the timing and duration of the period of time during which the simulation model predicted the occurence of stockouts was also qualitatively similar to the stockout periods observed empirically.

The second component of our validation relies on a survey of performance over the 4th quarter of 2009 that was independently conducted in 34 intermediate stocking HCs and 51 cross-docking HCs during the first quarter of 2010 as part of the public sector pilot evaluation protocol (see §2.2.3 for background and Friedman et al. [2011] for a full account of the pilot design and evaluation). Specifically, among other data the number of days of stockout (with a maximum possible value of 92) was measured from the stock cards of these HCs for all drugs in a tracer list that included the four pack sizes of AL. Our goal was to show that under similar operating conditions as those observed during the pilot, the simulation model would predict for these HCs a number of stockout days consistent with that measured empirically as part of this independent evaluation.

Because it was not clear how to aggregate stockout days across different pack sizes, we instanciated the model to simulate one pack size at a time using the ratios 2:1:1.5:2 for the demand of pack sizes 6, 12, 18 and 24 (respectively) that we had estimated from the stock cards of our dataset. By inspection of the stock cards we had collected through digital photography, we observed that most HCs received a large shipment of all four pack sizes simultaneously in May or June 2009 as part of the pilot preparation activities, so that we set June 2009 as the starting point of our simulations. Based on

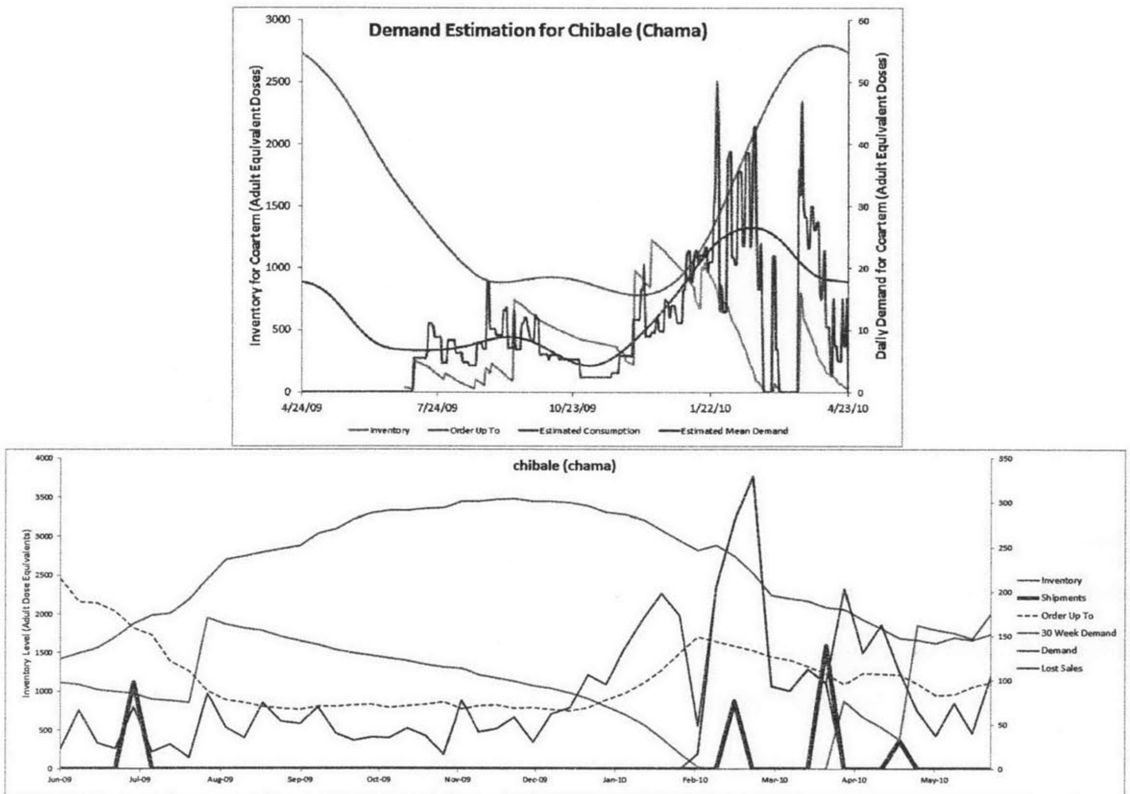


Figure 2-6: Trace of Inventory and Demand at Health Facility

Note. (Top) Estimated inventory and demand trace at Chibale HC, which was in a cross-docking district (Chama). (Bottom) Simulated inventory and demand trace at Chibale HC, under the cross-docking inventory policy.

the averages estimated over the subset of HCs for which we had detailed stock card data for either of these two months, we set the initial inventory levels for each pack size at approximately 80%, 170%, 120% and 100% of the maximum stock level specified by the policy. Finally, we simulated 50 replications from June to December 2009 for each pack size, using policy $IM_3^2[1, ami]$ in the 34 intermediate stocking HCs and policy $XM^4[1, ami]$ in the 51 cross-docking HCs, and estimated the number of days with a simulated stockout in each HC between October and December 2009[6]. We used the same discrete-event simulation dynamics, demand model and lead time model described in §§2.5.1, 2.5.1 and 2.5.1, respectively. Consistent with actual records of high inventory of AL available at MSL over that period of time, we also assumed unlimited availability of inventory at MSL in the simulation.

The results of these validation experiments along with the original survey counts of the number of stockout days are shown in Table 2.2. A key methodological remark is that given the inherent variability of this distribution system, the number of stockout days observed empirically should only be interpreted as the outcome of one sample path, not as the expectation of what the simulated means of stockout days over a large number of replications should be. Rather, the correct validation measure of proximity between actual and simulated results in this setting are the fractiles corresponding to the number of stockout days measured empirically relative to the simulated distribution of stockout days, which are shown in the last column of Table 2.2.

Table 2.2 thus shows that the actual measurements of stockout days are all fairly likely under the mathematical environment assumed for our simulation model. While the actual measurements of stockout days for pack sizes 6 and 18 in the cross-docking HCs seem slightly less likely (simulated fractiles of 0.96 and 0.08, respectively), we observe that the resulting difference in estimates remain relatively small in absolute terms. We suspect that this discrepancy is explained by the limitations of the stock card dataset that we used to estimate the initial inventory conditions for these specific pack

[6]To estimate the number of days without stock from the simulation output with a weekly time period, we assumed a constant daily demand within each week.

Policy	Pack Size	Pilot Mean	Simulation Mean	Percentile
A	6	12.97	13.55	0.44
A	12	5.06	4.11	0.78
A	18	7.56	7.26	0.62
A	24	8.09	10.20	0.20
B	6	3.16 (*)	1.34	0.96
B	12	0.00	0.09	0.56
B	18	0.20 (*)	0.80	0.08
B	24	1.80	1.24	0.84

Table 2.2: Empirical and Simulated Stock-Outs

Note. Pilot and simulation mean number of stockout days, which we denote by $\hat{N}_S$ and N_S respectively. The percentile is the percentile of the pilot mean $\hat{N}_S$ of the distribution of the simulation mean N_S. The symbol (*) indicates that we removed one HC from AL 6 and 18 for the cross-docking districts, because they were outliers, having an unusually high number of stockout days for that pack size compared to the other health centers.

sizes and HCs. On this basis we conclude that our simulation model seemingly offers a relatively realistic prediction of the service level achieved by an inventory distribution policy in this setting.

2.5.2 Results

The first set of simulation results shown in Figure 2-7 sheds light on the performance differentials between a version of our proposed inventory distribution policy XO and the inventory control policies XM and IM that are currently used in Zambia.

Focusing on the existing policies first, observe from Figure 2-7 (a) that $XM^4[1, ami]$ has a higher service level performance than $IM_3^2[1, ami]$ for all supply/demand ratios considered except the lowest. This is consistent with the field results summarized in Table 2.2 that were obtained independently as part of the public sector pilot, and is explained by the higher base stock level used by the cross-docking policy for HC replenishment (this is confirmed by the average HC inventory levels for these policies seen in Figure 2-7 (b)). While the rationale for testing $IM_3^2[1, ami]$ may have included

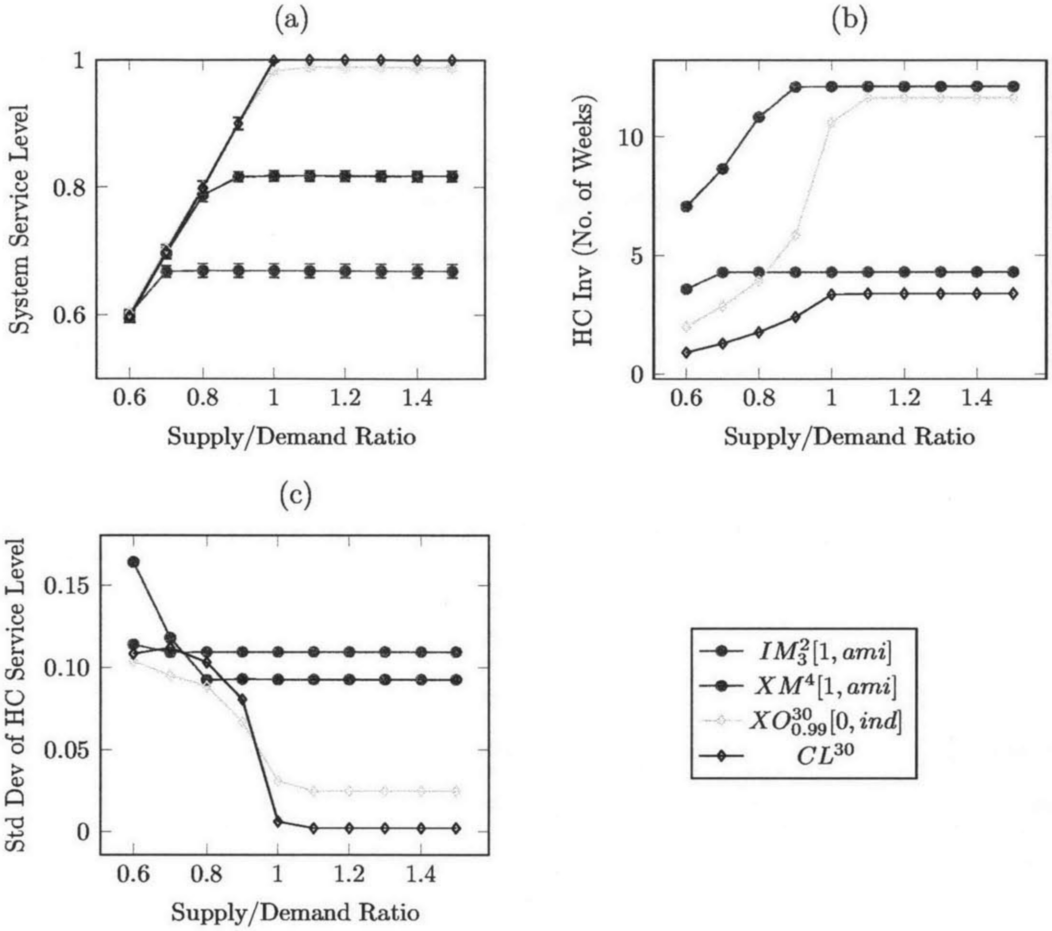


Figure 2-7: System Performance Metrics (Current and Optimization Policies)

the placement of more inventory at the DHOs which are closer to the patients than MSL, in practice secondary distribution lead times tend to dominate, and the addition by policies IM of specific receiving, picking and packing operations at the DHOs further increased these lead times. The results in Table 2.2, obtained for Q4 2009 under full supply conditions, may suggest on face value a service level of 96% or higher for the cross-docking policy $XM^4[1, ami]$, and may have been the basis for the decision made in late 2010 to progressively deploy that policy to the entire country. Consistent with the empirical results discussed in §2.3.2 that were obtained by analyzing stock cards over the entire year however, the average annual service level of $XM^4[1, ami]$ seen in Figure

2-7 remains only slightly above 80% without any noticeable increase observed for any values of the supply/demand ratio beyond 0.9. This observation is also aligned with the performance explanation provided in §2.3.2 that regardless of how much inventory is available upstream, current policies relying on average monthly issues over the last three months generate stockouts by ignoring seasonality and heterogeneity in both demand and lead times.

Figure 2-7 (a) also shows that policy $XO^{30}_{0.99}[0, ind]$ is able to close almost the entire gap in service level performance between the existing cross-docking policy $XM^4[1, ami]$ and the clairvoyant policy CL. Specifically, our proposed heuristic maintains with policies $XM^4[1, ami]$ and CL a service level equal to the supply/demand ratio for values of that ratio lower than 0.8 – note that the supply/demand ratio constitutes an upper bound on the average service level achievable by any policy, and that in situations of marked inventory scarcity achieving that bound only requires that no unreasonably high inventory be placed in any location (as $IM^2_3[1, ami]$ does by placing inventory at the DHO level for example). For supply/demand ratios higher than 0.8 however, policy $XO^{30}_{0.99}[0, ind]$ continues to achieve a service level only a couple percentage points below the upper bound (resulting in service levels of 97% and more for ratios greater than one), even though the min/max policy $XM^4[1, ami]$ stagnates at about 80% as discussed above. Because our proposed policy does not have access to the values of all future demand and lead times however, Figure 2-7 (b) shows that it does require higher average HC inventory levels to achieve this service performance than the clairvoyant policy (which by definition does not require safety stocks). Its average inventory level is still substantially lower than that of the existing policy $XM^4[1, ami]$ for supply/demand ratios lower than 1, but for higher values these two policies involve the same average HC inventory level of approximately 12 weeks of demand. Because the service level of our proposed policy is higher by almost 20% for high supply/demand ratios however, it implies that this same level of HC inventory is distributed by $XM^4[1, ami]$ across health center in a manner that does not closely match their respective demands. This is expected since that policy does not change its target replenishment levels between

HCs with different upcoming demand or lead times. Indeed, Figure 2-7 (c) shows that the standard deviation of service levels across HCs is substantially lower for our proposed policy than for the existing min/max policies, with the exception of the single supply/demand ratio value of 0.8 for which $XO^{30}_{0.99}[0, ind]$ and $XM^4[1, ami]$ perform similarly from the standpoint of access fairness.

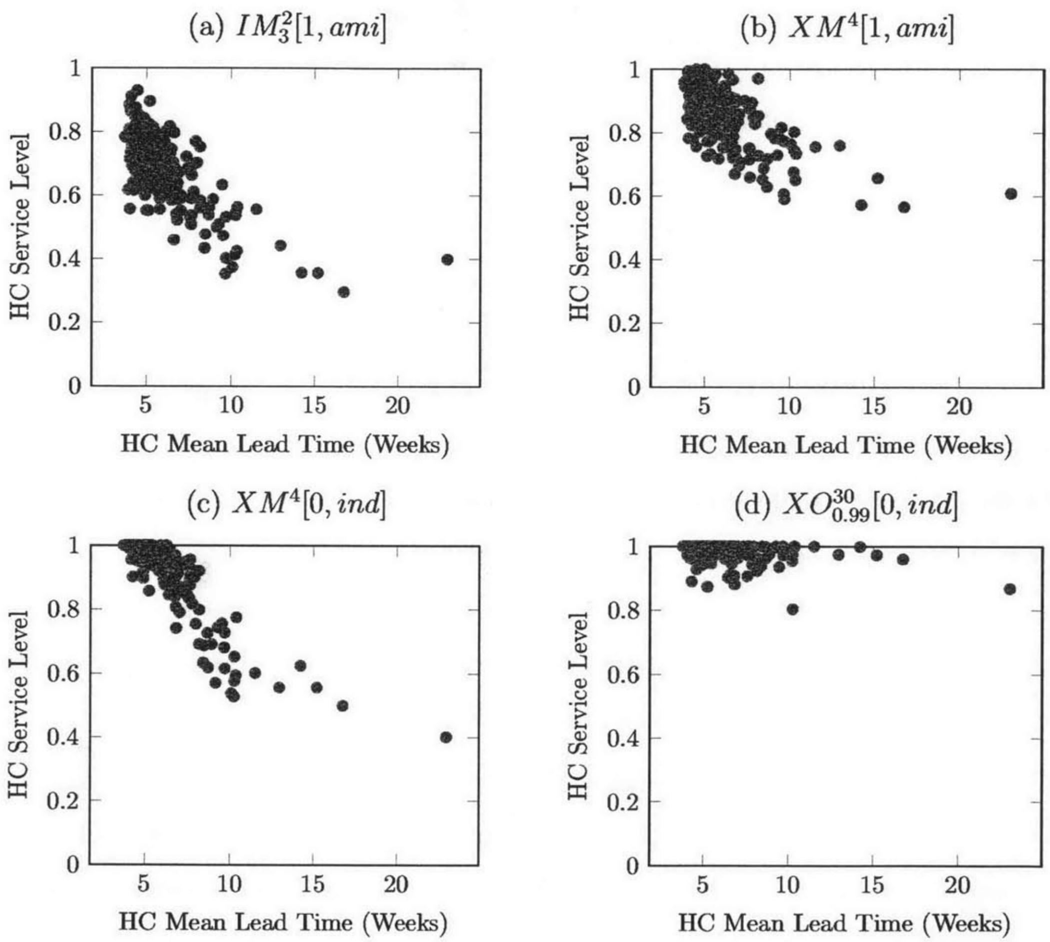


Figure 2-8: Plot of Service Level and Lead Time for Health Facilities

Figure 2-8 provides a representation of the variability in service levels associated with the four policies $IM^2_3[1, ami]$, $XM^4[1, ami]$, $XM^4[0, ind]$ and $XO^{30}_{0.99}[0, ind]$ that is arguably more intuitive than the measure of standard deviation used in Figure 2-

7 (c). These different representations are still consistent, as for example the vertical range of dots representing a single HC in Figure 2-8 (a), (b) and (d) (approximately [0.3, 0.95], [0.6, 1] and [0.8, 1] for policies $IM_3^2[1, ami]$, $XM^4[1, ami]$ and $XO_{0.99}^{30}[0, ind]$ respectively) is aligned with the corresponding standard deviation values in Figure 2-7 (c). But Figure 2-8 also confirms our interpretation that the more unfair access to medicines across HCs generated by the existing cross-docking policy is partly due to its blindness to differences in access lead times. Specifically, Figure 2-8 (a)–(c) exhibit a strong negative correlation between service level and mean access lead time. In contrast, Figure 2-8 (d) exhibits no such correlation, showing that in this case the differences of access lead times across HCs are properly accounted for by the optimization-based policy $XO_{0.99}^{30}[0, ind]$ (through constraint (2.6) in the formulation defining the heuristic). A comparison with Figure 2-8 (c) is particularly interesting, because the min/max policy $XM^4[0, ind]$ it relates to is provided with the exact same (instantaneous) communication infrastructure and high-quality demand forecasts as policy $XO_{0.99}^{30}[0, ind]$. In particular, its target replenishment levels rely on a forecast of demand over the *next* 4 months that is identical to that provided to our proposed policy. Policy $XM^4[0, ind]$ does not differentiate between HCs based on their access lead times however, so that Figure 2-8 (c) precisely shows that this lead time information is crucial to increase the fairness of access to medicines across different locations.

We next examine more systematically the role of communication delays and forecast quality on the performance of policies considered. Figure 2-9 shows plots of the three main performance metrics for different supply/demand ratios when the policies XM^4 and $XO_{0.99}^{30}$ are used in the two extreme infrastructure environments $[1, myo]$ and $[0, ind]$ (see §2.5.1 for definitions). The relatively small difference seen there between the performance of $XO_{0.99}^{30}[1, myo]$ and $XO_{0.99}^{30}[0, ind]$ suggests that our distribution heuristic is robust to increases in communication delays and a deterioration of forecast accuracy. This environmental change has a greater performance impact on the min/max policy XM^4 for the metrics of service level and average HC inventory however. We believe that the greater sensitivity of that policy to communication delays in

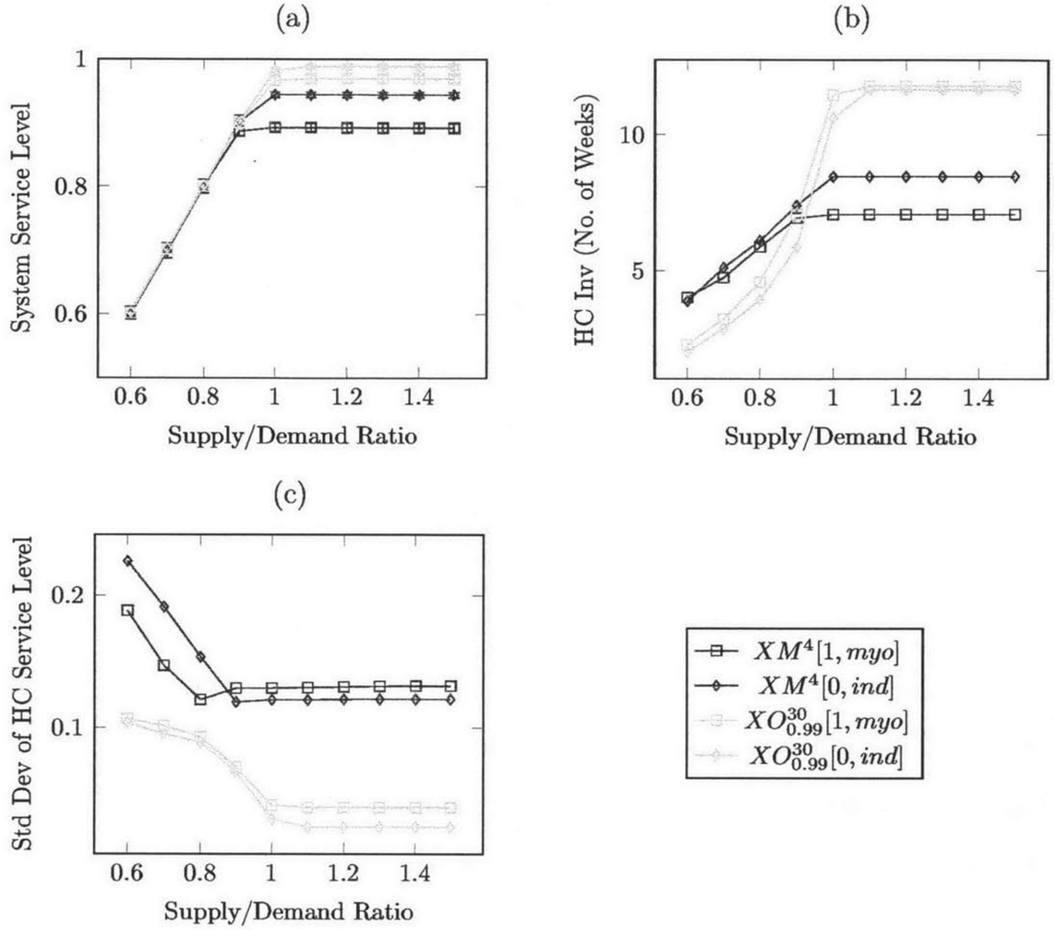


Figure 2-9: System Performance Metrics (Forecast-Based Policies)

particular is due to the potential mitigating role that faster information transmission may play on the blindness of XM^4 to differences in lead times between HCs. By comparison with Figure 2-7, Figure 2-9 also reveals that the performance increase of the min/max policy XM^4 along the dimensions of service level and inventory is substantial when going from the environment $[1, ami]$ to $[1, myo]$ (9pp service level) and $[1, myo]$ to $[0, ind]$ (6pp service level), suggesting that investments to develop forecasting capability for non-seasonal demand could be justified regardless of whether our proposed policy is eventually adopted. The service level of $XM^4[0, ind]$ remains lower to that of $XO^{30}_{0.99}[1, myo]$ by approximately three percentage points for all supply/demand ratios greater than 1 however, showing that for this metric even a much better infrastructure can not compensate for the coarse implicit assumption of stationary and homogeneous lead times. When considering the traditional trade-off of service level versus average inventory, the performance of $XM^4[0, ind]$ does seem quite good because as seen in Figure 2-9 (b) it does achieve an average inventory at the HCs which is 30% lower than that of $XO^{30}_{0.99}$. The high standard deviations of service levels across HCs seen in Figure 2-9 (c) does show however that this improved version of the min/max policy performs quite poorly along the dimension of access fairness across locations. That is, policy XM^4 achieves savings in average inventory level not by reducing shipments to all HCs equally, but rather by strongly penalizing the subset of locations which take longer to reach, as is confirmed by the results from Figure 2-8 (c) discussed earlier. That distribution fairness performance seems problematic both in absolute terms (range of service levels across locations $[0.4, 1]$ for a supply/demand ratio equal to 1.2) and relative to our proposed heuristic (range $[0.8, 1]$ for the same ratio).

Finally, we explore the impact of policy parameters on the distribution performance of these policies. Completing the set of experiments reported in Figures 2-7, Figure 2-10 displays the set of service level and average HC inventory performance measures achieved by the policies $XM^x[1, ami]$, $IM^{2x}_{3x}[1, ami]$, $XO^x_{0.99}[0, ind]$ and CL^x for different values of the supply/demand ratio when the main parameters x defining each of these policies spans a wide range of values. Similarly, Figure 2-11 completes Figure 2-9 by showing the policies $XM^x[0, ind]$, $XO^x_{0.99}[0, ind]$ and CL^x spanned for different values of the policy parameter x. We observe in Figures 2-10 and 2-11 that the efficient frontier of combined service and inventory performance achieved by $XO^x_{0.99}[0, ind]$ generally dominates that spanned by all the other min/max based policies. As discussed earlier, the superiority of our proposed distribution policy tends to decrease for low values of the supply/demand ratio, or even vanish altogether as seen in Figure 2-11 (c). This is not however a reflection of the policies, but rather follows from the fact that the inventory distribution problem becomes comparatively easier for low values of that ratio.

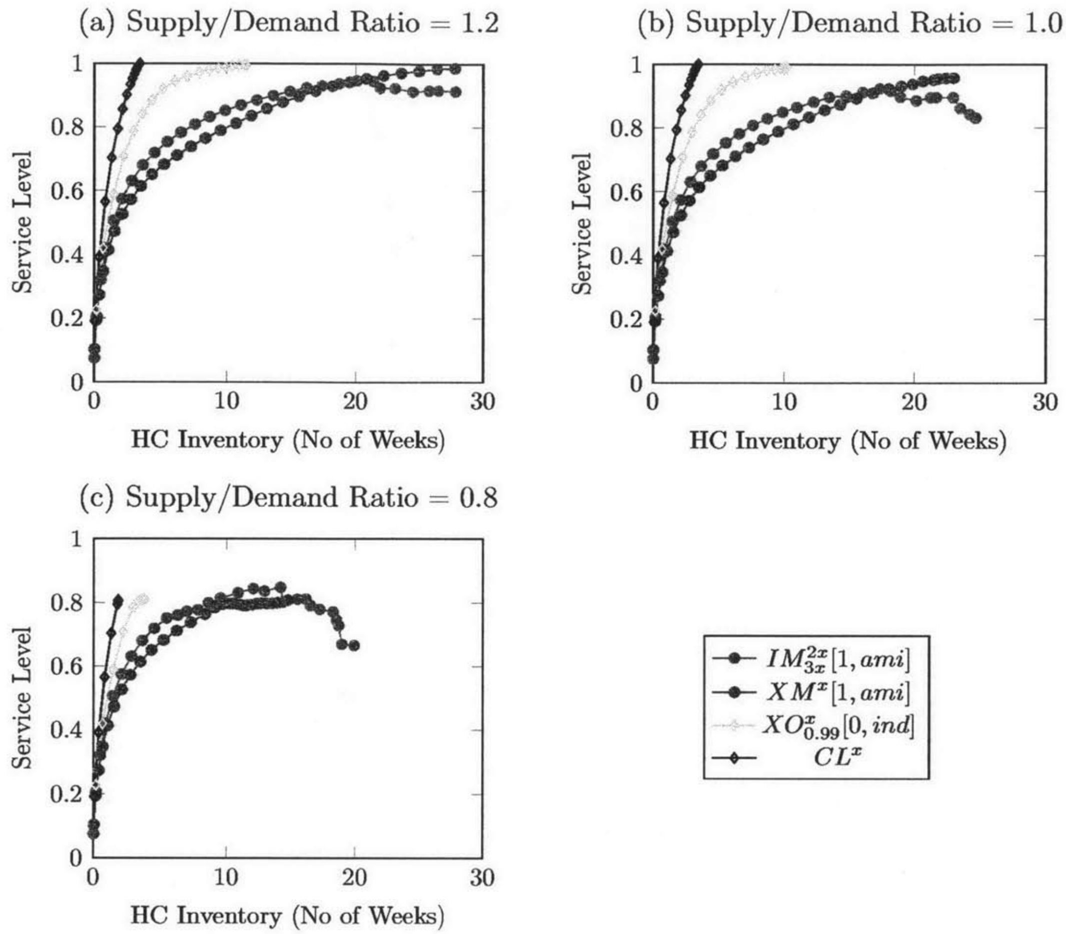


Figure 2-10: Service Level and Inventory Frontier (Current and Optimization Policies)

2.6 Conclusion

This paper focused on the technical aspects of the distribution of essential medicines in Zambia, against a backdrop of widespread stockouts of life-saving medicines and insufficient patient access to drugs in most of sub-Saharan Africa. A key goal of this work was to generate enduring knowledge on pharmaceutical distribution systems able to manage a large quantity of different health commodities in a scalable manner. For analysis purposes, we thus focused on the anti-malarial Arthemeter Lumefantrin (AL), which is particularly important to public health and combines a number of representa-

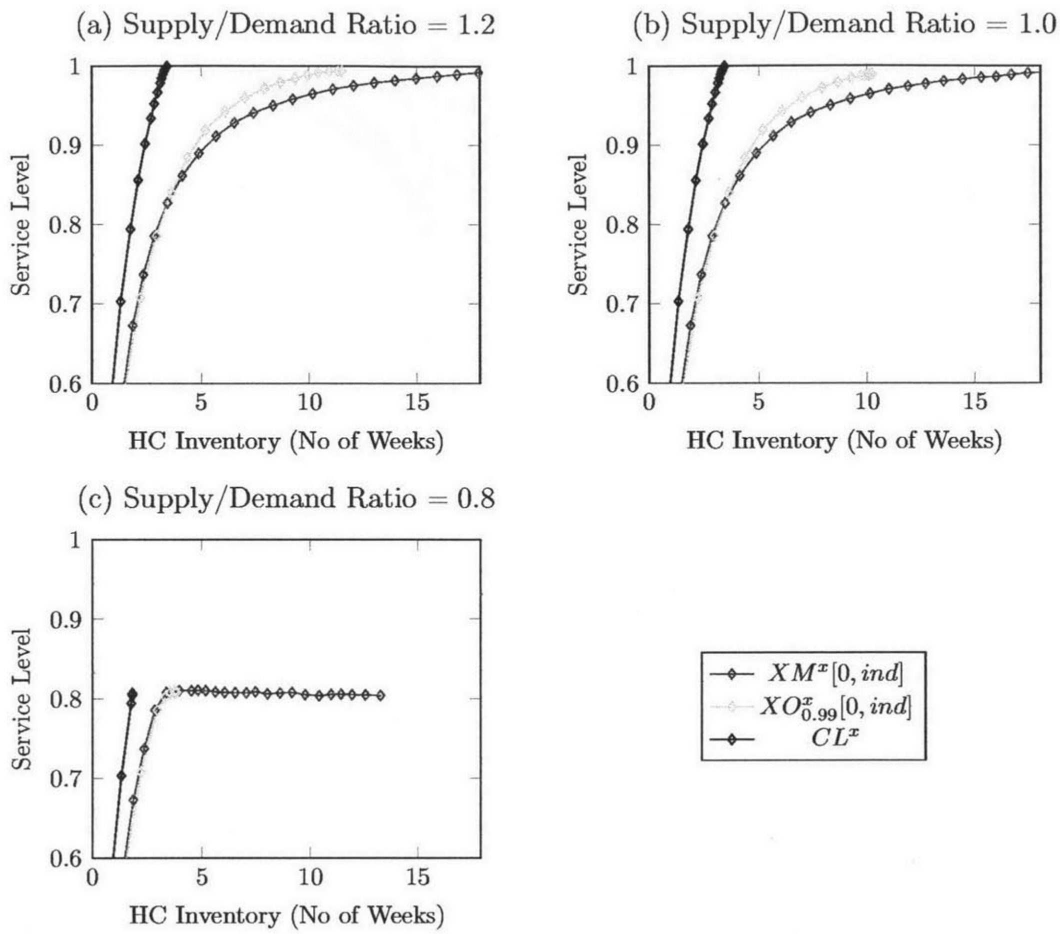


Figure 2-11: Service Level and Inventory Frontier (Forecast-Based Policies)

tive challenges because it is characterized by seasonal patient demand and needs to be distributed in a large number of locations with heterogeneous and seasonal access lead times.

To evaluate the performance of the inventory control system currently used in Zambia's public distribution system, we leveraged a landmark supply chain pilot experiment organized in 2009/2010 by the Ministry of Health with support from a number of international partners. Specifically, we constructed through digital photography a dataset of stock control cards providing continuous visibility on inventory and demand experienced in a number of patient-facing facilities through an entire year. From this dataset, we concluded that Zambia's current inventory control policies lead to predictable patient-level stock-outs, even when there is ample inventory available in the central warehouse, sufficient staff is available in local district pharmacies, and that staff properly adheres to prescribed inventory ordering procedures. Furthermore, we identified a clear explanation for these findings (the failure to properly anticipate seasonal variations in demand and supply lead-times), and were able to replicate them using simulation. This inventory control system happens to also be used in many other countries in Africa and beyond, and much public US money has been spent to promote and develop its use. These findings therefore seem significant for global health.

We shared and discussed in March of 2010 an early version of this paper with some of Zambia's key development partners in the area of pharmaceutical supply-chain management, including representatives of the DELIVER project which constitutes an important initiative funded by the US government in the area of pharmaceutical supply chain development aid. In March 2011, that organization issued a supplement [USAID | DELIVER, 2011b] to its earlier Logistics Handbook [USAID | DELIVER, 2011a] which specifically focused on the management of the supply chain for anti-malarials. We were pleased to read that the chapter of that publication dedicated to inventory control systems included recommendations such as "Before and during peak malaria periods, consider different ways to calculate resupply"; "Change how AMC is calculated"; "Use three months of consumption data from 12 months ago"; "Increase maximum stock level as lead time is longer"; and that its chapter on logistics management systems also discussed supply chain opportunities arising from mobile phones.

Much work remains in order to develop fully specified, sound and robust inventory distribution guidelines and disseminate them widely in sub-Saharan Africa however. To that end, we also develop in this paper a detailed proposal for an alternative inventory distribution system potentially relying on mobile devices and an associated inventory control heuristic. This heuristic is based on a simple mathematical optimization model inspired from existing systems already running successfully in other challenging distribution environments (e.g. Caro et al. [2010]). To evaluate this proposal, we leveraged the dataset previously mentioned in order to build a large-scale simulation model with validated predictive accuracy. Numerical experiments performed with that model suggest that the implementation of our proposed distribution policy would lead to substantial improvements of patient access to drugs relative to the current system. Our simulation results also suggest that a substantial performance increase of the existing inventory control policy would require the capabilities of reliable access lead times estimation and non-seasonal demand forecasting across the entire distribution network. These findings are significant from a practical standpoint, because the forecasting capability in particular is likely to require a more scalable and digital information system, which may be challenging to implement. Such a system would however also generate unprecedented opportunities for system transparency and accountability, reduction of paperwork linked to inventory management in chroniquely understaffed patient-facing facilities, and an evolution of demand estimation for procurement from notoriously unreliable epidemiology-based "quantification" exercises [Management Sciences for Health, 1997] to rigorous forecasting driven by actual demand data.

After the findings of an early version of this paper were presented to representatives of Zambia's Ministry of Health, Central Medical Stores, the Human Development and Research Groups of the World Bank and Crown Agents, decisions were made to fund and initiate an operational research project known as eZICS (enhanced inventory control system for Zambia) and designed to develop and evaluate a version of our proposed inventory distribution system as part of a controlled field pilot in the districts of Kassama, Kafue and Mkushi (about 100 health centers). A few weeks later, a partnership was formed with IBM in order to develop a field-ready and potentially scalable version of this system comprising deployed mobile phones with a client application providing ergonomic data entry capabilities for inventory transaction through a bar code scanner; a forecasting component with a user-friendly interface; a shipment optimization component interfacing with the legacy warehouse management software in place at the MSL warehouse in Lusaka; and a distributed transaction and performance reporting system for the entire region covered that is accessible through the internet. Under the technical leadership of a chief software architect from IBM's South Africa office who happens to be a Zambian national, a team of 6 software developers and database specialists from that firm has been working closely with us over the past 8 months in order to convert the models and research results presented here into a robust software system adapted to the requirements and demands of field use. Given development progress to date, we anticipate that the evaluation of this system in the field will begin later this year. We look forward to continuing our contribution to this initiative and to reporting the knowledge it will hopefully generate for improving pharmaceutical distribution systems and patient access to medicines in resource-limited countries.

Bibliography

S Axsäter, J Marklund, and E A Silver. Heuristic methods for centralized control of one-warehouse, n-retailer inventory systems. *Manufacturing Service Oper. Management*, 4(1):75–97, 2002.

J Barrington, Olympia Wereko-Brobby, Peter Ward, Winfred Mwafongo, and Seif Kungulwe. SMS for life: a pilot project to improve anti-malarial drug supply management in rural Tanzania using standard technology. *Malaria Journal*, 9:298, 2010.

A Cameron, M Ewen, D Ross-Degnan, D Ball, and R Laing. Medicine prices, availability, and affordability in 36 developing and middle-income countries: a secondary analysis. *The Lancet*, 373(9659):240–249, 2009.

Felipe Caro and Jeremie Gallien. Inventory management of a fast-fashion retail network. *Operations Research*, 58(2):257–273, 2010.

Felipe Caro, Jérémie Gallien, Miguel Díaz, Javier García, José Manuel Corredoira, Marcos Montes, José Antonio Ramos, and Juan Correa. Zara uses operations research to reengineer its global distribution process. *Interfaces*, 40(1):71–84, 2010.

John Foreman, Jérémie Gallien, Julie Alspaugh, Fernando Lopez, Rohit Bhatnagar, Chee Chong Teo, and Charles Dubois. Implementing supply routing optimization in a make-to-order manufacturing network. *Manufacturing & Service Operations Management*, 10(4):547–568, November 2010.

J Friedman, M Sjoblom, M Vledder, and P Yadav. Enhancing public supply chain management in Zambia. World Bank Working Paper, 2011.

L J Frost and M R Reich. *Access: How Do Good Health Technologies Get to Poor People in Poor Countries?* Cambridge: Harvard Center for Population and Development Studies, distributed by Harvard University Press, 2008.

A D Harries, E J Schouten, and S D Makombe. Ensuring uninterrupted supplies of antiretroviral drugs in resource-poor settings: an example from malawi. *Bulletin World Health Organization*, 85:152ï£¡–155, 2007.

D C Heath and P L Jackson. Modeling the evolution of demand forecasts with application to safety stock analysis in production/distribution systems. *IIE Transactions*, 26(3):17–30, 1994.

B B Kangwana, J Njogu, B Wasunna, S V Kedenge, and D N Memusi. Malaria drug shortages in Kenya: A major failure to provide access to effective treatment. *American Journal of Tropical Medicine and Hygiene*, 80:737ï£¡–738, 2009.

S Kraiselburd and P Yadav. Supply chains and global health: An imperative for bringing operations management scholarship into action. Production and Operations Management, in press, 2012.

S Lall and S Bibile. The political economy of controlling transnationals: the pharmaceutical industry in Sri Lanka, 1972–1976. *International Journal of Health Services*, 8(2):299–328, 1978.

R Levine, J Pickett, N Sekhri, and P Yadav. Demand forecasting for essential medical technologies. *American Journal of Law & Medicine*, 34:225–255, 2008.

S Makridakis, S Wheelwright, and R Hyndman. *Forecasting: methods and applications.* Wiley, 3rd ed edition, 1998.

M Mamdani and G Walker. Essential drugs in the developing world. *Health Policy and Planning*, 1:187–201, 1986.

Management Sciences for Health. *Managing Drug Supply: The Selection, Procurement, Distribution, and Use of Pharmaceuticals*. Kumarian Press, West Hartford, CT, 2nd edition edition, 1997.

A Pasquet, E Messou, D Gabillard, A Minga, A Depoulosky, S Deuffic-Burban, E Losina, K A Freedberg, C Danel, X Anglaret, and Y Yazdanpanah. Impact of drug stock-outs on death and retention to care among HIV-infected patients on combination antiretroviral therapy in Abidjan, Côte d'Ivoire. *PLoS One*, 5(10):e13414, 2010.

M Patel. Drug costs in developing countries and policies to reduce them. *World Development*, 11(3):195–204, 1983.

O F Picazo and F Zhao. Zambia health sector public expenditure review: accounting for resources to improve effective service coverage. World Bank Publications, 2009.

The PLoS Medicine Editors. Time for a "third wave" of malaria activism to tackle the drug stock-out crisis. *PLoS Medicine*, 6(11):e1000188, 2009.

E L Porteus. *Foundations of Stochastic Inventory Theory*. Stanford Business Books, 2002.

J Quick. Applying management science in developing countries: Abc analysis to plan public drug procurement. *Socio-Economic Planning Sciences*, 16(1):39–50, 1982.

M J Roberts, W Hsiao, P Berman P, and M R Reich. *Getting health reform right: a guide to improving performance and equity*. Oxford University Press, New York, 2004.

E J Schouten, A Jahn, and A Ben-Smith. Antiretroviral drug supply challenges in the era of scaling up ART in Malawi. *Journal of International AIDS Society*, 14(1):S4, 2011.

A Sieter. *A practical approach to pharmaceutical policy*. World Bank Publications, 2010.

Omar Sued, Casey Schreiber, Nora Girón, and Massimo Ghidinelli. HIV drug and supply stock-outs in Latin America. *The Lancet Infectious Diseases*, 11(11):810–811, 2011.

USAID | DELIVER. *The Logistics Handbook: A Practical Guide for the Supply Chain Management of Health Commodities*. USAID | DELIVER PROJECT, Task Order 1, Arlington, Virginia, second edition edition, 2011a. Accessed from http://deliver.jsi.com/dlvr_content/resources/allpubs/guidelines/LogiHand.pdf on September 6, 2012.

USAID | DELIVER. *Guidelines for Managing the Malaria Supply Chain: A Companion to the Logistics Handbook*. USAID | DELIVER PROJECT, Task Order 3, Arlington, Virginia, 2011b.

WHO Maximizing Positive Synergies Collaborative Group. An assessment of interactions between global health initiatives and country health systems. *The Lancet*, 373: 2137–2169, 2009.

Ricarda Windisch, Peter Waiswa, Florian Neuhann, Florian Scheibe, and Don de Savigny. Scaling up antiretroviral therapy in uganda: using supply chain management to appraise health systems strengthening. *Globalization and Health*, 7:25, 2011. doi: 10.1186/1744-8603-7-25.

World Health Organization. The world medicines situation, 2004.

World Health Organization. World malaria report 2010 summary, 2011.

P Yadav, H L Tata, and M Babaley. Supply chain management for essential medicines. In *World Medicines*. World Health Organization, 2011.

P Yadav, J Gallien, M Vledder, and M Sjoblom. Book chapter in preparation, 2012.

Prashant Yadav. Analysis of the public, private and mission sector supply chains for essential drugs in Zambia. Technical report, MIT-Zaragoza International Logistics Program, 2007.

J S Yudkin. Provision of medicines in a developing country. *The Lancet*, 311(8068): 810–812, 1978.

P Zipkin. Old and new methods for lost-sales inventory systems. *Operations Research*, 56:1256–1263, September/October 2008.

Paul H. Zipkin. *Foundations of Inventory Management*. McGraw Hill, 2000.

D Zurovac, M Ndhlovu, N Sipilanyambe, P Chanda, and DH Hamer. Paediatric malaria case-management with artemether-lumefantrine in zambia: a repeat cross-sectional study. *Malaria Journal*, 6:31, 2007.

D Zurovac, J K Tibenderana, J Nankabirwa, J Ssekitooleko, and J N Njogu. Malaria case-management under artemether-lumefantrine treatment policy in uganda. *Malaria Journal*, 7:181, 2008.

Chapter 3

Markdown Optimization for a Fashion Internet Retailer

3.1 Introduction

3.1.1 Motivation

Pricing is an essential business processes for a fashion retailer. A key reason is due to the fact that retail prices are highly visible to both customers and competitors, and have an immediate and dramatic impact on the bottom line. Ghemawat and Nueno [2003] estimate that an average fashion item is sold at 70% of its list price. If a fashion retailer were able to sell an average item at a larger percentage of the list price (e.g. increasing the average sales revenue to 75% of list price) by improving its pricing process, the retailer would be able to improve its profitability.

Indeed, since the 1990s, the fashion retail industry has seen a dramatic shift from manual, ad-hoc pricing practices towards data-driven analytical approaches that leverage sophisticated demand forecasting and optimization algorithms to recommend better pricing actions. This is evidenced by the formation of retail pricing software startups, many of which have been acquired by large ERP vendors, including SAS, SAP, JDA and Oracle. Retailers who have replaced their human pricing processes with software to manage their prices have seen "gains in gross margin in the range of 5% to 15%." [Friend and Walker, 2001]

Broadly speaking, there are two different categories of price adjustments in a retail setting: *promotions*—which are sales associated with holidays and events such as Mother's Day, Black Friday, or a Friends and Family event at a particular retailer; and *markdowns*—which are permanent discounts that occur during the life cycle of an item. For example, see Figure 3-1 for an illustration of a markdown price sequence. In practice, a markdown is often implemented as a reduction in the ticket price of an item, i.e. the retailer is required to replace the old price tags on the items with new price tags indicating the new ticket price. The focus of this work is the *markdown optimization* (MDO) problem, which aims at finding the timing and magnitude of markdowns to maximize the revenue obtained over the selling season of a seasonal or fashion item. Promotions are planned events with timing and discounts often fixed in advance of the start of the selling season, which can be easily incorporated as exceptions in the MDO problem. In practice, promotion planning and markdown planning are distinct processes at a retailer, with different objectives. The goal of promotions is to increase store traffic, and is typically a category-level decision (e.g. 20% off all sweaters); whereas markdowns are typically item-specific, and are more active typically towards the end-of-life of an item.

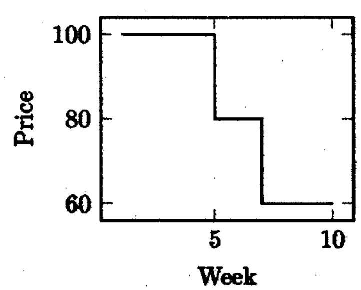


Figure 3-1: Markdown Price Sequence Example

One of the reasons why markdowns are ubiquitous in the fashion retail industry is that the retailer needs to reduce the price of the item in order to sell the current season's inventory and clear space for the next season's inventory. Another reason why markdown pricing is so common is that many fashion retailers use retail accounting, wherein the value of the inventory for accounting purposes is the number of units of the item multiplied by the current posted price of the item. The U.S. Securities and Exchange Commission does not allow retailers who practice retail accounting to increase the posted prices of their items after they have been marked down in order to prevent retailers from artificially inflating the value of their inventory and thus inflating the value of their company.

There has been significant work on the MDO problem in the retail industry, as well as in academia. To the best of our knowledge, the majority of the existing research has focused on the setting of a brick-and-mortar retailer. The focus of our research is to study the MDO problem in the context of an Internet retailer (e-retailer). A key feature which distinguishes online shopping is that an e-retailer is able to collect detailed information about customer behavior, such as customer patterns of arrivals or customer searches for related items. The goal of our research is to help e-retailers to use clickstream data in order to make better pricing decisions.

Online shopping is an important segment of the fashion retail industry, because online sales are growing rapidly, whereas brick-and-mortar sales growth is weak or negative. For example, in its first quarter of fiscal year 2015, online sales at the fashion retailer The Gap, Inc. increased 13.0% compared to the previous year, whereas same-store sales for the company's Gap brand were down 5.0% worldwide [Davis, 2014].

One of the reasons why customers prefer online shopping compared to shopping in brick-and-mortar stores is the convenience of shopping online. In developed countries, most people have Internet access and are able to visit an e-store using an Internet-connected device of their choosing, and have experience in shopping and buying products online. The time cost of visiting an e-store is negligible, as it takes less than a minute to visit an e-store using a device with an Internet connection; versus the time cost of travelling to and from a brick-and-mortar store. An e-store is also open at all times, whereas a brick-and-mortar store has restricted opening hours.

Because the experience of shopping online is so different from the experience of shopping at a brick-and-mortar store, we would expect customers who shop online to behave differently from customers who shop at brick-and-mortar stores. In particular, because online shopping is extremely convenient, we would expect customers to visit e-stores more frequently than they would visit brick-and-mortar stores; and we would also expect to observe a significant number of price-sensitive returning customers at e-stores, i.e. customers who do not buy the item during their initial visit, but return at a future time to buy the item at a lower price.

Indeed, we were able to find two different studies that suggest that a significant proportions of an e-retailer's customers are price-sensitive returning customers. The first study is a study of online buyer behavior by SeeWhy Inc. [Nicholls, 2014] which found that in their research sample, only 29% of the customers who add an item to their virtual shopping carts will purchase the item in the same session. The majority (71%) of customers will abandon their shopping carts, but 75% of the "abandoners" will return to the store in the future, either to purchase or to abandon again. SeeWhy also observed that in their research sample, returning visitors (which they defined as customers who had previously abandoned or purchased) made up only 12% of traffic, yet accounted for 36% of sales. The second study was one that we performed, based on data we obtained from a fashion Internet retailer. Our analysis showed that the average proportion of the total sales that are due to returning customers is 12.1%, and the average proportion of the total revenue that is due to returning customers is 19.4%. A detailed analysis of the data can be found in section 3.6.

The two studies suggest that in practice, a significant proportion of sales at an e-retailer are to returning customers. Our research is motivated by the question: can we formulate a markdown pricing model which takes into account the behavior of returning customers is order to make better pricing decisions which generate additional revenue?

One of the advantages of online shopping for retailers is that an e-retailer is able to use customer logins and/or cookies to collect a detailed record of the browsing and purchase behavior of customers. However, the fact that an e-retailer collects detailed customer data does not automatically ensure that the e-retailer will be able to use the data to make better operational decisions.

To the best of our knowledge, most of the pricing models that are used in practice for pricing fashion items in brick-and-mortar fashion retailers are based on demand models which are estimated using aggregate (e.g. weekly) sales data, because typically customer-level data is not available for brick-and-mortar stores. The goal of our research is to formulate a pricing model where the demand is estimated using customer-level data which can be collected by a fashion e-retailer, in order to make better pricing decisions. In particular, we believe that a pricing model that takes into consideration information about the behavior of returning price-sensitive customers will be able to earn additional revenue compared to a pricing model that only uses aggregate sales data.

3.1.2 Contributions

- *We propose a demand model for fashion e-retailers which models returning customer behavior, can be estimated from real data, which we incorporated into a pricing model.* Our proposed demand model can be estimated from clickstream data collected by a fashion Internet retailer.

- *We show that our proposed markdown pricing model is tractable.* For general non-linear first-time demand functions, the markdown pricing problem with returning customers is not concave. in the case when the first-time demand functions are linear, or in the case when the first-time demand arrives only in the first period and is exponential, we are able to show the existence and uniqueness of solutions to the markdown pricing model. We also propose a Jacobi iterative method for finding the optimal solution to the markdown pricing model, and prove a rate of convergence of the Jacobi iterative method.

- *We find conditions under which markdown prices are optimal for a dynamic pricing problem with returning customers, without imposing a markdown constraint.*

We found the following three cases when markdown prices are optimal for the dynamic pricing problem with returning customers:

1. The case when there are two periods, and the demand from first-time customers in period 1 is a multiple of the demand from first-time customers in period 2.
2. The case when the demand from first-time customers in each period is the same, and all the customers return in future periods.
3. The case of a returning demand model that has a declining sales property that is observed in practice.

- *We quantify the value of our markdown model.* The challenge is to consider a model where we can make a fair comparison of the returning pricing model and the myopic returning model. Furthermore, the markdown pricing model with returning customers cannot be solved in closed form. This motivates us to develop upper bounds on the ratio of the returning revenue relative to the myopic revenue (from traditional myopic markdown pricing models). We develop bounds for both the case of linear and nonlinear myopic demand functions.

- *We estimate our returning demand model and derive a pricing policy using actual data obtained from an e-retailer.* We obtained customer clickstream data from an actual fashion e-retailer. We compared the estimated myopic demand model and the estimated returning demand model, and showed that the returning demand model has comparable forecast accuracy with the myopic demand model. We predict higher revenues with the returning demand model using the optimized returning prices, compared to the myopic demand model using the optimized myopic prices.

3.1.3 Literature Review

This work is situated in the broader field of dynamic pricing. For an overview of the vast literature on dynamic pricing, we commend to the reader the excellent survey papers Elmaghraby and Keskinocak [2003], Bitran and Caldentey [2003], the book chapter Smith [2009], and the books Phillips [2005], Talluri and van Ryzin [2005]). Below, we give a brief overview of three streams of related research.

The first stream of related literature deals with the problem of dynamic pricing with strategic customers. The first paper to propose a game-theoretic model of a retailer selling to strategic customers is Coase [1972]. In this paper, Coase states his famous "Coase conjecture" that in the setting of a monopolist selling a durable good, if the monopolist is able to change prices instantaneously, then the monopolist will immediately price at his production cost. In practice, one can observe that the Coase conjecture does not hold for fashion e-retailers. Typically, a fashion e-retailer will sell an item over a sales horizon of 12–20 weeks, during which the item will take 3–6 retail prices. One of the most important reasons why the Coase conjecture does not hold for fashion e-retailers is because retailers are able to commit to a small number of price changes that are separated in time.

A subset of the literature on dynamic pricing with strategic customers focuses on the markdown optimization setting, which is the focus of our work. A non-exhaustive list of such papers which study markdown optimization problems as game-theoretic models include: Aviv and Pazgal [2008], Besanko and Winston [1990], Cachon and Swinney [2009], Correa et al. [2011], Osadchiy and Vulcano [2010], Liu and van Ryzin [2008]. Typically, these papers assume that each customer purchases at most one unit during the sales horizon, and that customers are rational and forward-looking, so that customers select the optimal time to purchase in order to maximize their expected individual surplus. The objective of the retailer is to set prices to maximize the retailer's expected (discounted) revenues. For the purpose of keeping the analysis tractable, these papers make certain assumptions about customer behavior. For example, Aviv and

Pazgal [2008] assumes that the sales horizon is divided into two periods, and all of the customers who decide not to buy in the first period remain in the store until the second period. While this stream of research is able to give useful managerial insight about how operational decisions such as prices or production quantities affect a fashion retailer's profitability, to the best of our knowledge, these models are not used operationally for making pricing decisions. In our work, we assume that customers are myopic in their purhcase behavior—i.e. each customer has a valuation for the item, and purchases the item during his first visit to the store in which the posted price is less than his valuation.

A second stream of research deals with the dynamic pricing problem faced by a retailer when customers make repeated purchases and their purchase behavior is affected by cognitive biases. In particular, based on exposure to recent past prices, customers form an internal expectation of the price which is referred to as the *reference price*. Prospect theory [Kahneman and Tversky, 1979] states that customers perceive the current retail price as a gain or loss relative to a reference price. Therefore, the current demand is larger (smaller) if the current price is smaller (larger) than the customers' reference price. The most common model for reference price formation in this literature is exponential smoothing [Kopalle et al., 1996, Fibich et al., 2003, Popescu and Wu, 2007]. Kopalle et al. [1996] and Fibich et al. [2003] study linear demand models with kinked linear reference effects, while Popescu and Wu [2007] study the more general case of nonlinear reference price effects. Unlike the aforementioned papers, Nasiry and Popescu [2011] studies the dynamic pricing problem under a different reference price formation process, where the reference price is a weighted average of the lowest and most recent prices. In a markdown pricing setting, prices decrease over time, and so a reference price formed using the "average" of past prices is not relevant because the current price will always be less than the "average" of the past prices. In our work, we assume that the most important factor driving intertemporal price effects are the returning behavior of customers. This is a different mechanism to model intertemporal price effects. Our model does have a similar effect with reference price effects is that in both models, if we hold the current price to be constant, higher (lower) past prices lead to higher (lower) current demand.

The final stream of related literature that we consider is related to the dynamic pricing problems with returning customers. Ahn et al. [2007] (henceforth AGK) considers a joint manufacturing/pricing decision problem, with linear demand and simple returning behavior.

The key distinguishing features of our work are the following: First, our demand model allows for nonlinear demand functions, and is able to model complex and realistic customer returning behavior (a customer may visit in non-consecutive periods e.g. visit at period 1, not return to visit in period 2, but return in period 3). Next, our work deals with the finite horizon markdown pricing problem. Lastly, as a proof of concept, we estimate our demand model using clickstream data obtained from a fashion e-retailer.

3.1.4 Structure of Paper

The rest of the paper is structured as follows. In section 2, we present the structure of a markdown optimization model without returning customers, and contrast this with our proposed markdown optimization model with returning customers (RMDO). In section 3, we prove the existence and uniqueness of a solution to the RMDO model. In section 4, we analyze properties of the optimal prices of our RMDO model. We first show that markdown prices are optimal for a dynamic pricing problem with returning customers, without imposing a markdown constraint in subsection 3.4.1, then study the relationship between returning and myopic prices in subsection 3.4.2. In section 5, we propose a model to quantify the value of the markdown model with returning customers. In section 6, we perform a numerical experiment using clickstream data obtained from an Internet retailer. We estimate both a myopic demand model and our returning demand model, and compare the forecast accuracy and the forecasted revenue improvement under myopic and returning markdown pricing. Section 7 concludes our findings.

3.2 Returning CustomerPricing Model

3.2.1 MDO Model without Returning Customers

Markdown optimization models which are used in practice for pricing at brick-and-mortar stores (see e.g. Smith [2009]) typically make the assumption that the demand during a time period is a function only of the posted price in that particular period. An equivalent way to interprete this assumption is to assume that the customers purchase behavior is described by the following assumption.

Assumption 1 (Non-returning customer behavior). Each customer has a valuation v for the item. Consider a given customer who arrives at time period $t \in \{1, 2, \dots, T\}$. If the customer's valuation exceeds the posted price p_t, then the customer will purchase the item (if inventory is available) and leave the store; otherwise the customer will leave the store, and do not return a future time period.

Under Assumption 1, the non-returning markdown optimization model has the following mathematical formulation:

$$\begin{aligned} \max_{p_t} \quad & \sum_{t=1}^{T} p_t f_t(p_t) \\ \text{s.t.} \quad & \sum_{t=1}^{T} f_t(p_t) \leq I \\ & p_1 \geq p_2 \geq \cdots \geq p_T \geq 0 \end{aligned} \tag{3.1}$$

where

p_t = the posted price of the item in period t.
I = the initial inventory level of the item.
f_t = the demand function for period t.

It is common in the revenue management literature to assume that the demand functions satisfy regularity conditions (see e.g. Talluri and van Ryzin [2005]), so that problem (3.1) is a well-behaved mathematical optimization problem. We assume that the demand functions of first-time customers satisfy the following conditions.

Let Ω_p denote the set of feasible prices.

Assumption 2 (Demand function). The first-time customer demand function is convex, non-negative, and strictly decreasing, i.e. $f_t'(p) < 0$, on Ω_p.

Assumption 3 (Revenue function is concave). The first-time customer revenue function $pf_t(p)$ is concave on Ω_p.

Assumption 4 (Finite maximum price). There exists a finite price p_∞ such that the first-time demand is zero at that price, i.e. $f_t(p_\infty) = 0$ for $t = 1, 2, \ldots, T$.

3.2.2 MDO Model with Returning Customers and Continuous Prices

One of the key assumptions that underlie the non-returning markdown optimization model is the assumption in Assumption 1 that customers who do not purchase the item leave the store and do not ever return in a future time period. We argued in subsection 3.1.1, that while this assumption applies in a brick-and-mortar store setting, we have evidence that to an e-retailer setting, there is a significant proportion of price-sensitive returning customers. Motivated by this finding, we propose to relax Assumption 1 to allow customers who do not purchase to return and possibly purchase in future time periods.

Assumption 5 (Returning customer behavior). Each customer has a valuation v for the item. Consider a given customer who arrives at time period $t \in \{1, 2, \ldots, T\}$. If the customer's valuation exceeds the posted price p_t, then the customer will purchase the item (if inventory is available) and leave the store; otherwise the customer will leave the store without purchasing the item, but may return in a future time period $u > t$ and purchase the item in period u if the posted price p_u is less than his valuation.

In Assumption 5, customers who do not purchase may return in future periods to purchase, but the probability distribution of their future return period is not specified.

In our demand model, we assume a return probability that depends on both the arrival period and the return period.

Assumption 6 (Return probability dependent on arrival and return periods). Consider a customer who visits the online store in period t but does not purchase the item. The probability that the customer returns in period $u > t$ is γ_{tu}. is independent of the posted price p_t and the customer's willingness to pay w.

Remark. In Assumption 6, we assume that the return probability is independent of the customer's valuation of the product, conditioned on the valuation being lower than the posted price in the arrival period. This assumption is reasonable because customers who would return in future periods are customers who are interested in the item, and so their valuation could not be too low. This implies that the range in valuation of returning customers is not large. Additionally, fashion items are not big-ticket items (e.g. a car or washing machine), for which we would expect customers to be more strategic in their purchase behavior due to the high cost of the item. We expect therefore that customers who did not purchase the item would return surrendipitously, rather than with a very high degree of intentionality.

Let $f_t(p_t)$ denote the demand induced by customers who arrive for the first time in period t. The demand in period t is the sum of the first-time demand in period t and the returning demand from previous periods, i.e.

$$d_t(p_1, \ldots, p_t) = f_t(p_t) + \sum_{u=1}^{t-1} r_{ut}$$

where r_{ut} is the returning demand from customers who visited the store but did not buy in period u and first return to the store in period t.

If all the customers who visited the item in period u were to return to visit the item in period t, then the demand due to those customers would be

$$\left[d_u(p_1, \ldots, p_{u-1}, p_t) - d_u(p_1, \ldots, p_{u-1}, p_u)\right]^+ .$$

(Here we use the notation $[x]^+ = \max\{x, 0\}$.) This is illustrated in Figure 3-2.

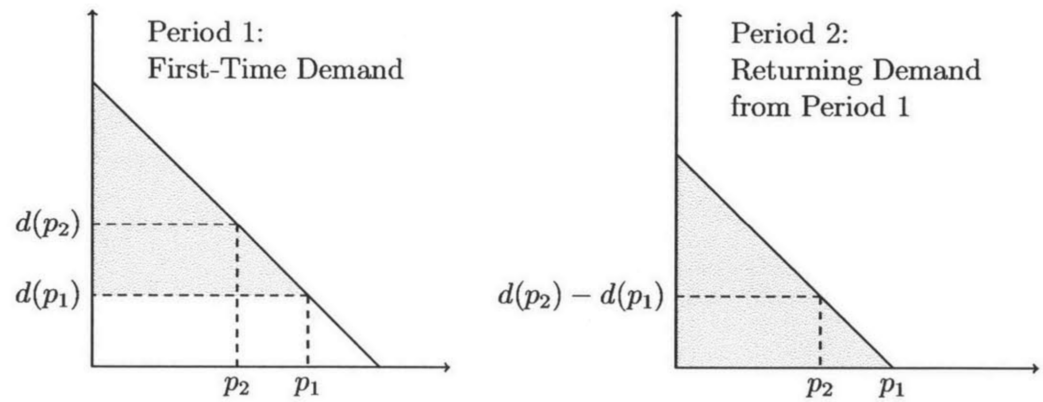


Figure 3-2: Illustration of Potential Returning Demand

However, since only the proportion γ_{ut} of these customers do return in period t, we have to multiply the demand by γ_{ut}, which gives

$$r_{ut} = \gamma_{ut}\left[d_u(p_1, \ldots, p_{u-1}, p_t) - d_u(p_1, \ldots, p_{u-1}, p_u)\right]^+ .$$

Thus we have the following expression for the demand in period t:

$$d_t(p_1, \ldots, p_t) = f_t(p_t) + \sum_{u=1}^{t-1} \gamma_{ut}\left[d_u(p_1, \ldots, p_{u-1}, p_t) - d_u(p_1, \ldots, p_{u-1}, p_u)\right]^+ . \qquad (3.2)$$

The markdown pricing problem formulation with returning customers is:

$$
\begin{aligned}
\max \quad & \sum_{t=1}^{T} p_t s_t \\
\text{s.t.} \quad & \sum_{t=1}^{T} s_t \leq I \\
& s_t = f_t(p_t) + \sum_{u=1}^{t-1} r_{ut} \\
& r_{ut} = [d_t(p_1, \ldots, p_{u-1}, p_t) - d_t(p_1, \ldots, p_{u-1}, p_u)] \\
& d_t(p_1, \ldots, p_t) = f_t(p_t) + \textstyle\sum_{u=1}^{t-1} \gamma_{ut} \left[d_u(p_1, \ldots, p_{u-1}, p_t) - d_u(p_1, \ldots, p_{u-1}, p_u)\right] \\
& p_1 \geq p_2 \geq \cdots \geq p_T \geq 0
\end{aligned}
\tag{RMDO}
$$

where

p_t	=	the posted price of the item in period t.
I	=	the initial inventory level of the item.
f_t	=	the demand function for customers who arrive for the first time in period t.
d_t	=	the total demand function for customers who visit the item in period t. In other words, the sum of the demand from customers who arrive for the first time in period t and from customers who did not buy in an earlier period but return in period t.
s_t	=	the demand which is equal to the sales in period t.
r_{tu}	=	the returning demand from customers who last visited the item in period t and first return to visit the item in period u

Remark. Due to the markdown constraint on prices in (RMDO), we always have $p_t \leq p_u$ in (3.2). Because the demand function $d_u(p_1, \ldots, p_{u-1}, \cdot)$ is a decreasing function, this implies that

$$[d_u(p_1, \ldots, p_{u-1}, p_t) - d_u(p_1, \ldots, p_{u-1}, p_u)]^+ = d_u(p_1, \ldots, p_{u-1}, p_t) - d_u(p_1, \ldots, p_{u-1}, p_u).$$

3.2.3 MDO Model with Returning Customers and Discrete Prices

In practice, many fashion retailers operate under a business constraint that the retail prices for all items must be chosen from a discrete feasible price set (e.g., {59.90, 49.90, 39.90}). The markdown optimization problem with returning customers and a discrete set of feasible prices can be formulated as an MIP optimization problem.

Before we state the formulation, we first need to introduce some notation. A customer who does not buy the item on his initial visit in period t_1 but returns to buy the item in period t_n may have returned to visit the store in the interim periods $t_2, \ldots, t_{n-1}$. Let us denote the set of all possible visit periods of returning customer visits by

$$\mathbf{V} = \{(t_1, \ldots, t_n) : n \geq 2, t_1 < \cdots < t_n, \{t_1, \cdots, t_n\} \subset \{1, \ldots, T\}\}.$$

Let us denote the probability that a customer would visit the store in the periods $(t_1, t_2, \ldots, t_n)$ by

$$\gamma_{t_1,\ldots,t_n} = \prod_{i=1}^{n-1} \gamma_{t_i t_{i+1}}.$$

Notation

- Discrete finite price set $\{q^1 > q^2 > \cdots > q^K\}$
- Set of time periods $\mathbb{T} = \{1, 2, \ldots, T\}$
- Set of price indices $\mathbb{K} = \{1, 2, \ldots, K\}$

Decision variables

- $\alpha_t^k = 1$ if the price q^k is selected during period t, i.e. $p_t = q^k$, and $\alpha_t^k = 0$ otherwise.
- $r_{t_1,\ldots,t_n}^k =$ demand realized in period t_n at the price q^k, from customers who visit the item during the periods $t_1, \ldots, t_n$, and purchase the item in period t_n. If the price q^k is not selected during period t_n, then $r_{t_1,\ldots,t_n}^k = 0$.

- s_t^k = sales in period t at price q^k. If the price q^k is not selected during period t, then $s_t^k = 0$.

$$
\begin{aligned}
\max \quad & \sum_{t=1}^{T}\sum_{k=1}^{K} q^k s_t^k & & (3.3)\\
\text{s.t.} \quad & \sum_{t=1}^{T}\sum_{k=1}^{K} s_t^k \le I & & (3.4)\\
& s_u^k \le f_u(q^k)\alpha_u^k + \sum_{\substack{(t_1,\dots,t_n):\mathbf{V}\\ t_n=u}} r_{t_1\dots t_n}^k & \forall u \in \mathbb{T} & \quad (3.5)\\
& r_{t_1\dots t_n}^k \le \gamma_{t_1,\dots,t_n}\sum_{j=1}^{k-1}[f_{t_1}(q^k) - f_{t_1}(q^j)]\alpha_{t_{n-1}}^j & \forall (t_1,\dots,t_n) \subset \mathbf{V} & \quad (3.6)\\
& r_{t_1\dots t_n}^k \le \gamma_{t_1,\dots,t_n}[f_{t_1}(q^k) - f_{t_1}(q^1)]\alpha_{t_n}^k & \forall (t_1,\dots,t_n) \subset \mathbf{V} & \quad (3.7)\\
& \sum_{k=1}^{K} q^k\alpha_t^k \ge \sum_{k=1}^{K} q^k\alpha_{t+1}^k & \forall t \in \{2,3,\dots,T\} & \quad (3.8)\\
& \sum_{k=1}^{K} \alpha_t^k = 1 & \forall t \in \mathbb{T} & \quad (3.9)\\
& \alpha_t^k \in \{0,1\} & \forall t \in \mathbb{T}, k \in \mathbf{K} & \quad (3.10)
\end{aligned}
$$

Constraint (3.5) sets the sales less than the total of the first-time demand and the returning demand from previous periods. Constraints (3.6) and (3.7) work together to set the correct upper bound on the returning demand in period t_n at the price q^k, from customers who visit in the periods $(t_1,\dots,t_n)$, i.e. $r_{t_1,\dots,t_n}^k$. Constraints (3.6) sets the upper bound on $r_{t_1,\dots,t_n}^k$ based on the lowest retail price the customers have seen before period t_n which is the price at period t_{n-1}. Constraint (3.7) sets $r_{t_1,\dots,t_n}^k$ to zero if the price q^k is not selected in period t, since the coefficient of $\alpha_{t_n}^k$ in (3.7) is the largest possible value of $r_{t_1,\dots,t_n}^k$. Constraint (3.8) is the markdown constraint $p_t \ge p_{t+1}$. Constraint (3.9) selects one of the prices q^k in period t, and constraint (3.10) is a binary constraint for the indicator variables.

The MIP formulation can be solved reasonably quickly in practice for problems of realistic size. The size of a representative practical MDO problem could have $T = 15$ periods and $K = 20$ prices in the price set. In theory, the number of variables $r_{t_1 \ldots t_n}$ can be exponential in T. However, in practice, the proportion of customers who visit over $n > 3$ periods (i.e. the return probabilities $\gamma_{t1 \ldots t_n}$) are usually insignificant. We can therefore solve a good approximation of the MIP with at most 3 visit periods by setting $\gamma_{t1 \ldots t_n}$ for $n > 3$. In this case, the number of decision variables and constraints is approximately $2T^3K \approx 135000$.

To illustrate correctness of the MIP formulation, consider the following example with $T = 3$ periods and $K = 3$ prices.

Example 3.2.1 ($T = 3$ periods and $K = 3$ prices).

$$
\begin{aligned}
&\max \sum_{t=1}^{T} \sum_{k=1}^{K} q^k s_t^k \\
&\text{s.t.} \sum_{t=1}^{T} \sum_{k=1}^{K} s_t^k \le I \\
&s_1^1 \le f_1(q^1)\alpha_1^1 \\
&s_1^2 \le f_1(q^2)\alpha_1^2 \\
&s_1^3 \le f_1(q^3)\alpha_1^3 \\
&s_2^1 \le f_2(q^1)\alpha_2^1 \\
&s_2^2 \le f_2(q^2)\alpha_2^2 + r_{12}^2 \\
&r_{12}^2 \le \gamma_{12}\big[f_1(q^2) - f_1(q^1)\big]\alpha_1^1 \\
&r_{12}^2 \le \gamma_{12}\big[f_1(q^2) - f_1(q^1)\big]\alpha_2^2 \\
&s_2^3 \le f_2(q^3)\alpha_2^3 + r_{12}^3 \\
&r_{12}^3 \le \gamma_{12}\Big(\big[f_1(q^3) - f_1(q^1)\big]\alpha_1^1 + \big[f_1(q^3) - f_1(q^2)\big]\alpha_1^2\Big) \\
&r_{12}^3 \le \gamma_{12}\big[f_1(q^3) - f_1(q^1)\big]\alpha_2^3 \\
&s_3^1 \le f_3(q^1)\alpha_3^1 \\
&s_3^2 \le f_3(q^2)\alpha_3^2 + r_{13}^2 + r_{23}^2 + r_{123}^2 \\
&r_{13}^2 \le \gamma_{13}\big[f_1(q^2) - f_1(q^1)\big]\alpha_1^1 \\
&r_{13}^2 \le \gamma_{13}\big[f_1(q^2) - f_1(q^1)\big]\alpha_3^2 \\
&r_{23}^2 \le \gamma_{23}\big[f_2(q^2) - f_2(q^1)\big]\alpha_2^1 \\
&r_{23}^2 \le \gamma_{23}\big[f_2(q^2) - f_2(q^1)\big]\alpha_3^2 \\
&r_{123}^2 \le \gamma_{12}\gamma_{23}\big[f_1(q^2) - f_1(q^1)\big]\alpha_2^1 \\
&r_{123}^2 \le \gamma_{12}\gamma_{23}\big[f_1(q^2) - f_1(q^1)\big]\alpha_3^2
\end{aligned}
$$

$$
\begin{aligned}
&s_3^3 \le f_3(q^3)\alpha_3^3 + r_{13}^3 + r_{23}^3 + r_{123}^3 \\
&r_{13}^3 \le \gamma_{13}\Big(\big[f_1(q^3) - f_1(q^1)\big]\alpha_1^1 + \big[f_1(q^3) - f_1(q^2)\big]\alpha_1^2\Big) \\
&r_{13}^3 \le \gamma_{13}\big[f_1(q^3) - f_1(q^1)\big]\alpha_3^3 \\
&r_{23}^3 \le \gamma_{23}\Big(\big[f_2(q^3) - f_2(q^1)\big]\alpha_2^1 + \big[f_2(q^3) - f_2(q^2)\big]\alpha_2^2\Big) \\
&r_{23}^3 \le \gamma_{23}\big[f_2(q^3) - f_2(q^1)\big]\alpha_3^3 \\
&r_{123}^2 \le \gamma_{12}\gamma_{23}\Big(\big[f_1(q^3) - f_2(q^1)\big]\alpha_2^1 + \big[f_1(q^3) - f_2(q^2)\big]\alpha_2^2\Big) \\
&r_{123}^2 \le \gamma_{12}\gamma_{23}\big[f_1(q^3) - f_1(q^1)\big]\alpha_3^3
\end{aligned}
$$

Let us illustrate how the constraints enforce the correct upper bounds for the demand in period t at price q^k. For example, if the price q^2 is selected in period 2, i.e. $\alpha_2^2 = 1$, then the demand in period 2 at price q^2 is equal to the sum of the first-time demand $f_2(q^2)$ and the returning demand from period 1 to 2 at the price q^2, r_{12}^2. The returning demand r_{12}^2 depends on the price in period 1. If $p_1 = q^1$, i.e. $\alpha_1^1 = 1$, then the returning demand $r_{12}^2 = \gamma_{12}[f_1(q^2) - f_1(q^1)]$; otherwise if $p_1 = q^2$, i.e. $\alpha_1^2 = 1$, then the returning demand $r_{12}^2 = 0$. We can check that the above inequalities enforce the correct upper bound for r_{12}^2. Also, if $\alpha_2^2 = 0$, we can see that the above inequalities enforce the constraint $s_2^2 = 0$.

We consider a fashion e-retailer selling a single item over a discrete time finite horizon of T periods. Because of long lead times for manufacture and shipment of fashion items, we assume that the e-retailer has an initial inventory level of I units of the item, and does not receive replenishment inventory. We assume without loss of generality that at the end of the sales horizon, the salvage value of the item is zero. The e-retailer's objective is to choose a non-increasing sequence of prices $p_1 \ge \cdots \ge p_T$ in order to maximize the revenue collected over the sales horizon. In this setting, the objective is typically revenue maximization because the cost of inventory is a sunk cost.

3.3 Existence and Uniqueness

Proposition 3.3.1. *A solution exists for the returning MDO problem* (RMDO).

Proof. The objective function is bounded, because the prices p_t are bounded above by a finite p_∞, and $\sum_{t=1}^{T} s_t$ is bounded above by I. The objective function is also continuous in the decision variables. Finally, the feasible region is bounded and closed. Therefore, by applying the extreme value theorem, the objective function attains its maximum. □

Let us define for convenience the revenue function

$$\text{Rev}(\mathbf{p}) = \sum_{t=1}^{T} p_t d_t(p_1, \dots, p_t).$$

Consider the function g which we define by

$$g(\mathbf{p}) = \begin{bmatrix} g_1(\mathbf{p}) \\ \vdots \\ g_T(\mathbf{p}) \end{bmatrix},$$

where $g_t(\mathbf{q})$ is the solution to the optimization problem

$$g_t(\mathbf{q}) = \underset{q_{t+1} \le p \le q_{t-1}}{\arg\max}\ \text{Rev}(q_1, \dots, q_{t-1}, p, q_{t+1}, \dots, q_T).$$

Intuitively, each function g_t is the best response function which is the price p_t^* that optimizes the revenue given that the other prices are fixed to $p_u = q_u$ for $u = 1, \dots, T$ and $u \neq t$.

Applying the function g repeatedly is similar to a Jacobi iterative method for finding the solution of a linear system of equations.

We are able to show that when the first-time demand functions are elements of two classes of demand functions—linear and exponential—that are commonly used in practice [Talluri and van Ryzin, 2005], the returning MDO problem (RMDO) has a unique solution.

Theorem 3.3.2 (Uniqueness for linear first-time demand functions). *If the first-time demand functions are linear, i.e. $f_t(p_t) = a_t - b_t p_t$, then the returning markdown optimization model has a unique solution. Furthermore, the Jacobi iterative method converges to the unique solution with convergence rate*

$$\frac{\left\|\mathbf{N}^T(\mathbf{p}-\mathbf{p}^*)\right\|_1}{\|\mathbf{p}-\mathbf{p}^*\|_1} \leq 1 - \frac{1}{2^T}.$$

The proof of the theorem relies on the following definition and lemma.

Definition. An $n \times n$ real matrix $\mathbf{A}$ is said to be a *linear returning matrix* if the following conditions hold:

$$\begin{aligned} \sum_{i=1}^{j-1} a_{ij} &\leq \frac{1}{2} && j = 2, 3, \ldots, n \\ \sum_{i=j+1}^{n} a_{ij} &\leq \frac{1}{2} && j = 1, 2, \ldots, n-1 \\ a_{ii} &= 0 && i = 1, 2, \ldots, n \end{aligned}$$

Lemma 3.3.3. *If $\mathbf{A}$ is a $n \times n$ linear returning matrix, then $\|A^n \mathbf{x}\|_1 \leq (1 - 1/2^n)\|\mathbf{x}\|_1$.*

Proof. First, we show that $\|\mathbf{A}\mathbf{x}\|_1 \leq \|\mathbf{x}\|_1$. This can be seen by noting that

$$\|\mathbf{A}\mathbf{x}\|_1 \leq \sum_{i=1}^{n} x_i \sum_{j=1}^{n} a_{ji} \leq \sum_{i=1}^{n} x_i = \|\mathbf{x}\|_1. \tag{3.11}$$

Next, we show that the statement $S(k) : \left\|\mathbf{A}^k \mathbf{e}_k\right\|_1 \leq 1 - 1/2^k$ is true for $k = 1, 2, \ldots, n$ by induction on k. Here $\mathbf{e}_k$ denotes the k-th unit vector, which has a 1 in the k-th coordinate and value 0 in the other coordinates.

The base case $S(1)$ is true because

$$\|\mathbf{A}\mathbf{e}_1\|_1 = \sum_{i=1}^{n} a_{i1} \leq \frac{1}{2}.$$

Suppose that $S(k)$ is true for $1 \leq k < n$, we now show that $S(k+1)$ is true. Let $\mathbf{v} = \mathbf{A}\mathbf{e}_{k+1}$. We have

$$\left\|\mathbf{A}^{k+1}\mathbf{e}_{k+1}\right\|_1 = \left\|\mathbf{A}^k\mathbf{v}\right\|_1 \leq \sum_{i=1}^{k} v_i\left\|\mathbf{A}^k\mathbf{e}_i\right\| + \sum_{i=k+2}^{n} v_i\left\|\mathbf{A}^k\mathbf{e}_i\right\|.$$

We first observe that from the definition of the linear returning matrix,

$$\sum_{i=1}^{k} v_i \leq \frac{1}{2},$$
$$\sum_{i=k+2}^{n} v_i \leq \frac{1}{2}.$$

For $i = 1, 2, \ldots, k$, we can bound

$$\left\|\mathbf{A}^k\mathbf{e}_i\right\| = \left\|\mathbf{A}^{k-i}\mathbf{A}^i\mathbf{e}_i\right\| \leq \left\|\mathbf{A}^i\mathbf{e}_i\right\| \leq 1 - \frac{1}{2^i}.$$

Note that we used (3.11) in the first inequality, and the induction statement $S(i)$ in the second inequality. This implies that

$$\sum_{i=1}^{k} v_i\left\|\mathbf{A}^k\mathbf{e}_i\right\| \leq \frac{1}{2} - \frac{1}{2^{k+1}}.$$

Using the bound

$$\sum_{i=k+2}^{n} v_i\left\|\mathbf{A}^k\mathbf{e}_i\right\| \leq \sum_{i=k+2}^{n} v_i \leq \frac{1}{2}$$

gives us the proof of $S(k+1)$.

Finally, we prove the proposition by

$$\left\|\mathbf{A}^n\mathbf{x}\right\|_1 \leq \sum_{i=1}^{n} x_i\left\|\mathbf{A}^n\mathbf{e}_i\right\|_1 \leq \left(1 - \frac{1}{2^n}\right)\sum_{i=1}^{n} x_i = \left(1 - \frac{1}{2^n}\right)\left\|\mathbf{x}\right\|_1.$$

$\square$

We now prove Theorem 3.3.2.

Proof of Theorem 3.3.2. The objective function can be rewritten as

$$\sum_{t=1}^{T} p_t \left(a_t - b_t p_t + \sum_{u=1}^{t-1} \gamma_{ut} c_u (p_u - p_t) \right),$$

where

$$c_t = b_t + \sum_{u=1}^{t-1} \gamma_{ut} c_u.$$

By setting the first-derivative with respect to p_t equal to zero, we get the optimality conditions $\mathbf{M}\mathbf{p} = \mathbf{a}$, where

$$m_{tu} = \begin{cases} -\gamma_{tu} c_t & \text{if } t < u \\ 2c_u & \text{if } t = u \\ -\gamma_{ut} c_u & \text{if } t > u \end{cases}$$

The Jacobi iteration is then $\mathbf{p}^{k+1} = \mathbf{A}\mathbf{p}^k + \mathbf{c}$ where

$$a_{tu} = \begin{cases} -\dfrac{\gamma_{tu} c_t}{2c_u} & \text{if } t < u \\ 0 & \text{if } t = u \\ -\frac{\gamma_{ut}}{2} & \text{if } t > u \end{cases}$$

We observe that the matrix $\mathbf{A}$ is a linear returning matrix because

$$\sum_{t=1}^{u-1} |a_{ut}| \le \frac{\sum_{t=1}^{u-1} \gamma_{tu} c_t}{2c_u} \le \frac{1}{2} \qquad \forall u = 2, 3, \ldots, T,$$

$$\sum_{t=u+1}^{T} |a_{tu}| \le \sum_{t=u+1}^{T} \frac{\gamma_{ut}}{2} \le \frac{1}{2} \qquad \forall u = 1, 2, \ldots, T-1.$$

Applying Lemma 3.3.3 proves the theorem. □

We show that for $T = 2$ periods and exponential first-time demand in period 1 with no first-time demand in period 2 that the returning MDO problem (RMDO) has a unique solution.

Theorem 3.3.4 (Uniqueness for exponential first-time demand functions). *Suppose that the first-time demand function in period* 1 *is exponential, i.e.* $f_1(p) = a_1 e^{-b_1 p}$, *and there is no first-time demand in period* 2, *i.e.* $f_2(p) = 0$. *Then the returning MDO problem* (RMDO) *has a unique solution. Furthermore, the Jacobi iterative method converges to the unique solution with convergence rate*

$$\frac{\|g^2(\mathbf{p}) - \mathbf{p}^*\|_1}{\|\mathbf{p} - \mathbf{p}^*\|_1} \leq \frac{\gamma}{2}.$$

Proof. We prove the theorem by showing that an optimal solution to (RMDO) must satisfy the fixed-point equation $g(p_1, p_2) = [p_1,\ p_2]$. By showing that g^2 is a contraction, the Banach fixed point-theorem implies that there is a unique solution to the fixed-point equation and therefore a unique solution to (RMDO).

For a two-period problem, there is only a single return probability γ_{12}, so we will denote $\gamma = \gamma_{12}$ for convenience.

The objective function is

$$a_1 p_1 e^{-b_1 p_1} + \gamma a_1 p_2 [e^{-b_1 p_2} - e^{-b_1 p_1}].$$

We first determine $g_1(p_2)$. Setting the partial derivative of the objective with respect to p_1 to zero gives us

$$e^{-b_1 p_1} - b_1 p_1 e^{-b_1 p_1} + \gamma b_1 p_2 e^{-b_1 p_1} = 0.$$

Therefore

$$g_1(p_2) = \frac{1}{b_1} + \gamma p_2.$$

We next determine $g_2(p_1)$. Setting the partial derivative of the objective with respect to p_2 to zero gives us

$$e^{-b_1 p_2} - e^{-b_1 p_1} - b_1 p_2 e^{-b_1 p_2} = 0.$$

We can manipulate this to

$$p_1 = p_2 - \frac{\log(1 - b_1 p_2)}{b_1}.$$

This implies that

$$g_2(p_1) = h^{-1}(p_1), \qquad \text{where } h(p) = p - \frac{\log(1 - b_1 p)}{b_1}.$$

We are able to bound $(h^{-1})'(p_1) \leq 1/2$ by deducing that

$$h'(p) = 1 + \frac{1}{1 - b_1 p},$$

and that for $p_2 \in [0, 1/b_1)$, we have $h'(p_2) \geq 2$.

We have

$$g^2(p_1, p_2) = g\left(\frac{1}{b_1} + \gamma p_2, h^{-1}(p_2)\right) = \left[\frac{1}{b_1} + \gamma h^{-1}(p_1) \quad h^{-1}\left(\frac{1}{b_1} + \gamma p_2\right)\right]^T.$$

Therefore

$$\begin{aligned}
&\left\|g^2(p_1, p_2) - g^2(q_1, q_2)\right\|_1 \\
&= \gamma\left|h^{-1}(p_1) - h^{-1}(q_1)\right| + \left|h^{-1}\left(\frac{1}{b_1} + \gamma p_2\right) - h^{-1}\left(\frac{1}{b_1} + \gamma q_2\right)\right| \\
&\leq \frac{\gamma}{2}|p_1 - q_1| + \frac{\gamma}{2}|p_2 - q_2| \\
&= \frac{\gamma}{2}\|(p_1, p_2) - (q_1, q_2)\|_1.
\end{aligned}$$

This shows that g^2 is a contraction with the Lipschitz constant $\gamma/2$, and proves the theorem. $\square$

3.4 Pricing Policy Structure

3.4.1 When Markdowns are Optimal

In this section, we consider the dynamic pricing problem with infinite inventory and returning customers:

$$\max_{p_t} \sum_{t=1}^{T} p_t d_t(p_1, \ldots, p_t) \tag{3.12}$$

We analyze three settings under which the optimal prices of (3.12) are markdown prices, i.e. the markdown constraint is automatically satisfied.

Setting 1: Two Periods

Proposition 3.4.1. *If $T = 2$, and the period 1 first-time demand function is a multiple of the period 2 first-time demand function, i.e. $f_2(p) = cf_1(p)$ for some $c > 0$, then the optimal prices for the dynamic pricing problem* (3.12) *are markdown prices.*

Proof. In this case, the revenue function is

$$\text{Rev}(p_1, p_2) = p_1 f_1(p_1) + cp_2 f_2(p_2) + p_2\gamma\left[f_1(p_2) - f_1(p_1)\right]^+.$$

This is a special case of the returning demand function defined in (3.2).

Suppose by contradiction that the optimal prices are (p_1, p_2) with $p_1 < p_2$. This implies that there is no returning demand in period 2, i.e.

$$\text{Rev}(p_1, p_2) = p_1 f(p_1) + cp_2 f(p_2).$$

We will find another set of markdown prices (q_1, q_2) such that $\text{Rev}(q_1, q_2) > \text{Rev}(p_1, p_2)$, which contradicts the optimality of (p_1, p_2).

Consider the ordering of $p_1 f(p_1)$ and $p_2 f(p_2)$. There are three possible cases.

Case 1: $p_1 f(p_1) < p_2 f(p_2)$ In this case, we have

$$\text{Rev}(p_2, p_2) = (c+1)p_2 f(p_2) > \text{Rev}(p_1, p_2) = p_1 f(p_1) + cp_2 f(p_2).$$

Case 2: $p_1 f(p_1) > p_2 f(p_2)$ In this case, we have $\text{Rev}(p_1, p_1) > \text{Rev}(p_1, p_2)$.

$$\text{Rev}(p_1, p_1) = (c+1)p_1 f(p_1) > \text{Rev}(p_1, p_2) = p_1 f(p_1) + cp_2 f(p_2).$$

Case 3: $p_1 f(p_1) = p_2 f(p_2)$ In this case, we can show that $\text{Rev}(p_2, p_1) > \text{Rev}(p_1, p_2)$ because

$$\begin{aligned}
\text{Rev}(p_1, p_2) &= p_1 f(p_1) + p_2 c f(p_2) \\
&= (c+1)p_1 f(p_1) \\
\text{Rev}(q_1, q_2) &= p_2 f(p_2) + p_1 c f(p_1) + p_1 \gamma [f(p_1) - f(p_2)] \\
&= (c+1)p_1 f(p_1) + p_1 \gamma [f(p_1) - f(p_2)].
\end{aligned}$$

Therefore,

$$\text{Rev}(p_1, p_2) - \text{Rev}(q_1, q_2) = p_1 \gamma [f(p_1) - f(p_2)] \geq 0.$$

This is because $p_1 \geq 0, \gamma \geq 0$, and $p_1 < p_2 \implies f(p_1) > f(p_2)$ by Assumption 2. □

Setting 2: All Customers are Returning

Proposition 3.4.2. *Suppose that inventory is infinite; the first-time customers in each period are identical, i.e.* $f_t = f$*; and all customers are returning, i.e.* $\gamma_{t,t+1} = 1$*. Let* $\mathbf{p} = (p_1, p_2, \ldots, p_T)$ *be a markdown price sequence. Let* $\mathbf{q}$ *be any other price sequence formed by permuting the prices in* $\mathbf{p}$*. In other words, given a permutation* $\sigma : \{1, \ldots, T\} \to \{1, \ldots, T\}$*, we define* $q_{\sigma(t)} = p_t$ *for all* t*. Then*

$$\text{Rev}(\mathbf{p}) \geq \text{Rev}(\mathbf{q}),$$

which implies that the price sequence $\mathbf{q}$ *is not optimal.*

Proof. For convenience, we define the revenue function from the customers who arrive for the first time in period t, given that the price sequence is $\mathbf{q}$, by

$$\mathrm{Rev}_v(\mathbf{q}) = q_v f(q_v) + \sum_{t=v+1}^{T} q_t \left[f(q_t) - f\left(\min_{\tau=v,\dots,t-1} \{q_\tau\} \right) \right]^+ .$$

By definition,

$$\mathrm{Rev}(\mathbf{p}) = \sum_{t=1}^{T} \mathrm{Rev}_t(\mathbf{p})$$
$$\mathrm{Rev}(\mathbf{p}) = \sum_{t=1}^{T} \mathrm{Rev}_t(\mathbf{q})$$

We prove the result by showing that

$$\forall u = 1, \dots, T: \qquad \mathrm{Rev}_u(\mathbf{p}) \geq \mathrm{Rev}_v(\mathbf{q}), \qquad \text{where } v = \sigma(u). \tag{3.13}$$

Since $\mathbf{p}$ is a markdown price vector, therefore

$$\mathrm{Rev}_u(\mathbf{p}) = p_u f(p_u) + \sum_{t=u+1}^{T} p_t [f(p_t) - f(p_{t-1})]. \tag{3.14}$$

Let $\mathbf{V}$ denote the set of time indices $t \geq v+1$ such that q_t is the lowest price since period v, which is defined mathematically as

$$\mathbf{V} = \left\{ t \in \{v+1, \dots, T\} \;\middle|\; q_t < \min_{\tau=v,\dots,t-1} q_\tau \right\}.$$

Let $\{v_1 < v_2 < \cdots < v_N\}$ denote the elements of $\mathbf{V}$. By definition, the subsequence $\{q_{v_1}, q_{v_2}, \dots, q_{v_N}\}$ is a markdown price vector. Therefore we can write

$$\mathrm{Rev}_v(\mathbf{q}) = q_v f(q_v) + \sum_{n=1}^{N} q_{v_n} [f(q_{v_n}) - f(q_{v_{n-1}})],$$

where we define $v_0 = v$ for convenience.

Now by defining $u_n = \sigma^{-1}(v_n)$ for $n = 1, \ldots, N$, we can write $\mathrm{Rev}_v(\mathbf{q})$ in terms of prices p_t as

$$\mathrm{Rev}_v(\mathbf{q}) = p_u f(p_u) + \sum_{n=1}^{N} p_{u_n}[f(p_{u_n}) - f(p_{u_{n-1}})], \tag{3.15}$$

where we define $u_0 = u$ for convenience. Note that $\{p_{u_1}, \ldots, p_{u_N}\}$ is a subsequence of the vector $\{p_{u+1}, \ldots, p_T\}$.

To prove the desired inequality (3.13), we note that for $t = u, \ldots, T-1$, the coefficient of $f(p_t) - f(p_{t+1})$ is p_{t+1} in (3.14), but is $p_{u_{n+1}}$ where $n+1$ is the smallest number such that $u_n \geq t+1$ in (3.15). □

Setting 3: Declining Sales Model

We introduce a new demand model, which we refer to as the *declining sales model*, that is a special case of the returning demand model.

Our motivation for the declining sales model is the observation made in Heching et al. [2002] that the deseasonalized sales of a fashion item "decline with the number of weeks since the [markdown] price was implemented or the style was introduced." The declining sales model is based on the following assumptions about customer behavior:

1. There is a finite population of customers, whose associated demand function is $f(p)$. If all the customers visit the item in the store in period 1, then the demand function would be $f(p)$.

2. The customer visit behavior is described by the visit probability parameter $\alpha < 1$. In each period, each customer visits the item in the store with probability α. More rigorously, in each period, each customer conducts an i.i.d. Bernoulli trial with success probability equal to α, and visits the item in the store if and only if the Bernoulli trial is a success.

Based on these assumptions, we can derive that the first-time demand functions are

$$f_t(p) = \alpha(1-\alpha)^{t-1} f(p),$$

and the return probabilities are

$$\gamma_{tu} = \alpha(1-\alpha)^{u-t}.$$

We can also show that the demand in the declining sales model decline with the multiplicative factor $(1-\alpha)$ from the first period, and after each markdown.

Proposition 3.4.3. *Consider the declining sales model with visit probability $\alpha < 1$. For any markdown price sequence $\mathbf{p} = (p_1, \ldots, p_T)$, the demand declines with the multiplicative factor $(1-\alpha)$ from the first period, and after each price change.*

Proof. We can define any markdown price sequence $\mathbf{p} = (p_1, \ldots, p_T)$, by a decreasing sequence of prices $q_1 > q_2 > \cdots > q_N$, and an increasing sequence of time periods $1 = u_1 < u_2 < \cdots < u_N < u_{N+1} = T+1$, where $p_t = q_n$ if $u_n \le t < u_{n+1}$. Intuitively, the pair (q_n, u_n) indicates that at the time period u_n, the price is marked down to q_n.

Consider the price changes at the periods u_n and u_{n+1}. Our goal is to show that the demand sequence between two price changes forms a decreasing geomtric progression with common ratio $(1-\alpha)$.

We can divide the population of customers at the beginning of time period u_n into n different groups, based on the lowest price that a customer has observed. For $i = 1, \ldots, n-1$, let group i consist of the customers who have visited the item in the past, and the lowest price that the customer has observed is q_i. What is important to note is that by definition, customers in group i whose willingness to pay is greater than q_i have already bought the product before period u_n. We can also define group 0 to be the customers who have never visited the item, and for convenience, we can describe these customers as having "observed" the price $q_0 = +\infty$. For each group $0, 1, \ldots, n-1$, we can let π_i be the proportion of the total population which belongs in group i.

We can then write

$$d_t(p_1,\dots,p_t) = \alpha(1-\alpha)^{t-u_n} \sum_{i=0}^{n-1} \pi_i \left[f(q_n) - f(q_i)\right], \qquad t = u_n, u_n+1, \dots, u_{n+1}-1.$$

The reason for this expression is a customer in group i who has not already bought the product will buy the product during the customer's first visit to the item in the periods $u_n, u_n+1, \dots, u_{n+1}-1$. The probability that such a customer's first visit is in period t is given by the probability mass function from the geometric distribution, $\alpha(1-\alpha)^{t-u_n}$. □

Theorem 3.4.4. *Consider the declining sales model with visit probability α. Suppose that inventory is infinite. Then any optimal price vector is a markdown price vector.*

To prove Theorem 3.4.4, we need to use the following lemma.

Lemma 3.4.5. *Consider the declining sales model with visit probability α. Suppose that inventory is infinite. Suppose that $\mathbf{p} = (p_1, \dots, p_T)$ is a price vector such that $p_1 \geq p_2 \geq \cdots \geq p_u$ but $p_u < p_{u+1}$. Let us define the price vector*

$$\mathbf{q} = (p_1, \dots, p_{u-1}, p_{u+1}, p_u, p_{u+2}, \dots, p_T),$$

which is the same prices as $\mathbf{p}$ except that we have switched the prices in periods u and $u+1$. Then $\mathrm{Rev}(\mathbf{q}) \geq \mathrm{Rev}(\mathbf{p})$.

Proof. We can partition the set of customers into four non-overlapping sets:

1. Group $\mathbf{C}_{u,u+1}$ visits the item in both periods u and $u+1$.

2. Group $\mathbf{C}_u$ visits the item in period u but not in period $u+1$.

3. Group $\mathbf{C}_{u+1}$ visits the item in period $u+1$ but not in period u.

4. Group $\mathbf{C}_\varnothing$ does not visit the item in periods u nor $u+1$.

Let us denote by $\text{Rev}_{\mathbb{C}}(\mathbf{p})$ the revenue from the price vector $\mathbf{p}$ of the customers in the set $\mathbb{C}$. We can then write $\text{Rev}(\mathbf{p})$ and $\text{Rev}(\mathbf{q})$ as the sum of the revenue from four groups of customers:

$$\begin{aligned}\text{Rev}(\mathbf{p}) &= \text{Rev}_{\mathbf{C}_\varnothing}(\mathbf{p}) + \text{Rev}_{\mathbf{C}_u}(\mathbf{p}) \\ &\quad + \text{Rev}_{\mathbf{C}_{u+1}}(\mathbf{p}) + \text{Rev}_{\mathbf{C}_{u,u+1}}(\mathbf{p}) \\ \text{Rev}(\mathbf{q}) &= \text{Rev}_{\mathbf{C}_\varnothing}(\mathbf{q}) + \text{Rev}_{\mathbf{C}_u}(\mathbf{q}) \\ &\quad + \text{Rev}_{\mathbf{C}_{u+1}}(\mathbf{q}) + \text{Rev}_{\mathbf{C}_{u,u+1}}(\mathbf{q})\end{aligned}$$

We note that

$$\text{Rev}_{\mathbf{C}_\varnothing}(\mathbf{p}) = \text{Rev}_{\mathbf{C}_\varnothing}(\mathbf{q})$$

as the customers in $\mathbf{C}_\varnothing$ see the exact sequence of prices, since the prices $\mathbf{p}$ and $\mathbf{q}$ are identical except for periods u and $u+1$.

We note that

$$\text{Rev}_{\mathbf{C}_u}(\mathbf{p}) = \text{Rev}_{\mathbf{C}_{u+1}}(\mathbf{q})$$

as the customers see the exact sequence of prices, since the prices $p_u = q_{u+1}$. Similarly, we note that

$$\text{Rev}_{\mathbf{C}_{u+1}}(\mathbf{p}) = \text{Rev}_{\mathbf{C}_u}(\mathbf{q})$$

as the customers see the exact sequence of prices, since the prices $p_{u+1} = q_u$.

Finally, we note that

$$\text{Rev}_{\mathbf{C}_{u,u+1}}(\mathbf{p}) \leq \text{Rev}_{\mathbf{C}_{u,u+1}}(\mathbf{q}).$$

To prove this, notice that for the price vector $\mathbf{p}$, customers who have a willingness to pay greater than or equal to p_{u+1} and who did not buy the product before period u, will buy the product in period u at the price p_u; whereas for the price vector $\mathbf{q}$, customers who have a willingness to pay greater than or equal to p_{u+1} and who did not buy the product before period u, will buy the product in period u at the price $p_{u+1} > p_u$. For all other customers, they would buy the product at the same price for both price vectors $\mathbf{p}$ and $\mathbf{q}$. By Assumption 2, there is a positive mass of customers whose willingness to pay is in $[p_{u+1}, +\infty)$. This implies that we have a strict inequality

$$\text{Rev}_{\mathbf{C}_{u,u+1}}(\mathbf{p}) < \text{Rev}_{\mathbf{C}_{u,u+1}}(\mathbf{q}).$$

□

We are now ready to prove Theorem 3.4.4.

Proof of Theorem 3.4.4. We will prove this result by contradiction. Suppose that there exists an optimal price sequence $\mathbf{p}$ which is not a markdown price sequence. Then there exists a u such that $p_u < p_{u+1}$. By applying the interchange argument in Lemma 3.4.5, we get a price sequence $\mathbf{q}$ with strictly better revenue, i.e. $\text{Rev}(\mathbf{q}) > \text{Rev}(\mathbf{p})$. This is a contradiction of the optimality of $\mathbf{p}$. □

3.4.2 Returning Prices vs Myopic Prices

In this section, we analyze the relationship between the optimal returning prices and the optimal prices from a myopic pricing model, which does not take into account the impact of the current pricing decision on future revenue. In particular, we wish to investigate the impact of a change in the return probabilities—e.g. due to marketing efforts by the e-retailer—on the relative ordering of returning and myopic prices. We show that when inventory is infinite, the returning prices are always higher than the myopic price in the corresponding period, regardless of the returning probabilities; but this may not be true when inventory is finite.

The returning prices, p_t^r, are defined as the solution to the MDO model with returning customers (RMDO).

The *myopic prices* for the MDO model with returning customers (RMDO) and infinite inventory are defined as

$$p_1^m = \arg\max_{0 \le p \le p_\infty} p f_1(p),$$

and for $t = 2, 3, \ldots, T$

$$p_t^m = \arg\max_{0 \le p \le p_{t-1}} p f_t(p) + \sum_{u=1}^{t-1} \gamma_{ut} \left[d_u(p_1^m, \ldots, p_{u-1}^m, p_t) - d_u(p_1^m, \ldots, p_{u-1}^m, p_u^m) \right].$$

In the case when inventory is infinite, then we show that the returning price is higher than the myopic price in the same period.

Theorem 3.4.6. *Suppose that inventory is infinite, and that all of the return probabilities are positive, i.e.* $\gamma_{tu} > 0$. *Then the optimal returning price is greater or equal to the corresponding myopic price in the same period, i.e.* $p_t^r \ge p_t^m$.

The proof of the Theorem 3.4.6 relies on the follow increasing property of the optimal prices.

Proposition 3.4.7. *Consider the RMDO problem with* T *periods. For any two time periods* u, v *with* $1 \le u < v \le T$, *with the prices* p_t *fixed for* $t \in \{1, \ldots, T\} \setminus \{u, v\}$, *the* p_v *that optimizes the total revenue is an increasing function of* p_u.

Proof. Let us define the optimal price p_v as a function of p_u

$$h(p_u) = \arg\max_{p_{v+1} \le p \le p_{v-1}} \mathrm{Rev}(p_1, \ldots, p_{v-1}, p, p_{v+1}, \ldots, p_T).$$

The terms in the revenue function $\mathrm{Rev}(p_1, \ldots, p_{v-1}, p, p_{v+1}, \ldots, p_T)$ that depend on p_v are

$$p_v f_v(p_v) + p_v \sum_{t=1}^{v-1} \gamma_{tv} \left[d_t(p_1, \ldots, p_{t-1}, p_v) - d_t(p_1, \ldots, p_{t-1}, p_t) \right] \\ + \sum_{t=v+1}^{T} \gamma_{vt} p_t \left[d_v(p_1, \ldots, p_{v-1}, p_t) - d_v(p_1, \ldots, p_{v-1}, p_v) \right].$$

We can thereofre write

$$h(p_u) = \underset{p_{v+1} \le p \le p_{v-1}}{\arg\max} \left\{ \phi(p_v) + \gamma_{uv} p_v \left[d_u(p_1, \ldots, p_{u-1}, p_v) - d_u(p_1, \ldots, p_{u-1}, p_u) \right] \right\},$$

where $\phi(p_v)$ is a concave function of p_v for $p_v \le p_{v-1}$, and does not depend on p_u. Let us define the function ψ_{p_u}

$$\psi_{p_u}(p_v) := \phi(p_v) + \gamma_{uv} p_v \left[d_u(p_1, \ldots, p_{u-1}, p_v) - d_u(p_1, \ldots, p_{u-1}, p_u) \right].$$

Notice that for $\epsilon > 0$, we have

$$\psi'_{p_u+\epsilon}(p_{v-1}) - \psi'_{p_u+\epsilon}(p_{v-1}) = \gamma_{uv} \left[d_u(p_1, \ldots, p_{u-1}, p_u) - d_u(p_1, \ldots, p_{u-1}, p_u + \epsilon) \right] > 0. \tag{3.16}$$

In the case when $\psi'_{p_u}(p_{v-1}) \ge 0$, then because ψ_{p_u} is concave, that implies that $\psi_{p_u}(p_v)$ is increasing at the point $p_v = p_{v-1}$. This implies that $h(p_u) = p_{v-1}$. Due to (3.16), we also have that $\psi'_{p_u+\epsilon}(p_{v-1}) \ge 0$ which implies that $h(p_u + \epsilon) = p_{v-1}$.

In the case when $\psi'_{p_u}(p_{v-1}) < 0$, then because ψ_{p_u} is concave, that implies that the maximum of the function ψ_{p_u} is attained at a point $p_v < p_{v-1}$. Due to (3.16), we also have that $\psi'_{p_u+\epsilon}(p_v) < 0$ which implies that $h(p_u + \epsilon) > h(p_u) = p_v$. □

Proof of Theorem 3.4.6. We will prove by induction that $p_t^r \ge p_t^m$.

The base case is that $p_1^r \ge p_1^m$. We prove by contradiction that $p_1^r \ge p_1^m$. Suppose that $p_1^r < p_1^m$. Let us define the markdown price vector $\mathbf{q} = (p_1^m, p_2^r, p_3^r, \ldots, p_T^r)$. In words, the price vector $\mathbf{q}$ sets the myopic price for the first period and the returning prices for periods $2, 3, \ldots, T$. We will show that $\text{Rev}(\mathbf{p}^r) < \text{Rev}(\mathbf{q})$, which contradicts the assumption that Since the revenue function $pf_1(p)$ is concave by assumption, and $p_1^r < p_1^m$, therefore

$$p_1^r f_1(p_1^r) < p_1^m f_1(p_1^m).$$

We also have

$$p_t^r d_t(\mathbf{q}) - p_t^r d_t(\mathbf{p}^r) = p_t^r \gamma_{1t} \left[f_1(p_1^r) - f_1(p_1^m) \right] > 0.$$

We conclude therefore that

$$\mathrm{Rev}(\mathbf{q}) = p_1^m f_1(p_1^m) + \sum_{t=2}^{T} p_t^r d_t(\mathbf{q}) \geq p_1^r f_1(p_1^r) + \sum_{t=2}^{T} p_t^r d_t(\mathbf{q}) = \mathrm{Rev}(\mathbf{p}^r).$$

The inductive case is that for some $t \in \{1, 2, \ldots, T-1\}$, we have shown that $p_1^r \geq p_1^m$, $\ldots$, $p_t^r \geq p_t^m$. We prove by contradiction that $p_{t+1}^r \geq p_{t+1}^m$. Suppose that $p_{t+1}^r < p_{t+1}^m$. Let us define the markdown price vector $\mathbf{q} = (p_1^r, \ldots, p_t^r, p_{t+1}^m, p_{t+2}^r, \ldots, p_T^r)$. In words, the price vector $\mathbf{q}$ sets the myopic price for first period $t+1$ and the returning prices for all other periods. Using Proposition 3.4.7, we observe that

$$\begin{aligned}
p_{t+1}^m &= \arg\max_{p \leq p_t^m} p d_{t+1}(p_1^m, p_2^m, p_3^m, \ldots, p_t^m, p) \\
&\leq \arg\max_{p \leq p_t^m} p d_{t+1}(p_1^r, p_2^m, p_3^m, \ldots, p_t^m, p) \\
&\leq \arg\max_{p \leq p_t^m} p d_{t+1}(p_1^r, p_2^r, p_3^m, \ldots, p_t^m, p) \\
&\vdots \\
&\leq \arg\max_{p \leq p_t^r} p d_{t+1}(p_1^r, p_2^r, p_3^l, \ldots, p_t^r, p)
\end{aligned}$$

Given that $pd_{t+1}(p_1^r, p_2^r, p_3^l, \ldots, p_t^r, p)$ is a concave function of p, and that

$$p_{t+1}^r \leq p_{t+1}^m \leq \arg\max_{p \leq p_t^r} p d_{t+1}(p_1^r, p_2^r, p_3^l, \ldots, p_t^r, p),$$

this implies that

$$p_{t+1}^r d_{t+1}(\mathbf{p}^r) \leq p_{t+1}^m d_{t+1}(\mathbf{q}).$$

We also have for $u = t+2, \ldots, T$ that

$$p_u^r d_u(\mathbf{q}) - p_u^r d_u(\mathbf{p}^r) = p_u^r \gamma_{t+1,u} \left[f_{t+1}(p_{t+1}^r) - f_{t+1}(p_{t+1}^m) \right] > 0.$$

We therefore conclude that $\mathrm{Rev}(\mathbf{q}) > \mathrm{Rev}(\mathbf{p}^r)$ which contradicts our assumption that $\mathbf{p}^r$ is optimal. □

In the case when inventory is finite, we have a counterexample that shows that the conjecture that the returning price is higher than the myopic price in the same period, $p_t^r \geq p_t^m$, is not true for all $t = 1, 2, \ldots, T$. We can however show that $p_1^r \geq p_1^m$ even when inventory is finite, by using the same logic in the proof of Theorem 3.4.6 of $p_1^r \geq p_1^m$.

Example 3.4.8 (Returning price is not always higher than myopic price). In this example, we have $T = 2$ periods.

When inventory is finite, the myopic prices are defined inductively, i.e. p_1^m optimizes the revenue in period 1, p_2^m optimizes the revenue in period 2 given $p_1 = p_2^m$, and so on. In the case of $T = 2$ periods, we have

$$p_1^m = \arg\max_{0 \leq p_1} p_1 \min\{I, f_1(p_1)\},$$

$$p_2^m = \arg\max_{0 \leq p_2 \leq p_1^m} p_2 \min\{\max\{0, I - f_1(p_1^m)\}, f_2(p_2) + \gamma_{12}(f_1(p_2) - f_1(p_1^m))\}.$$

In this example, suppose that $f_1(p) = f_2(p) = 1 - p$, and that $I = 1$. It can be shown that for all γ, $p_1^m = 0.5, p_2^m = 0.5$. We also were able to determine numerically the optimal returning prices p_2^r as a function of γ. The returning and myopic prices are shown in Figure 3-3. While the period 1 returning price is higher than the period 1 myopic price, the period 2 returning price is always lower than the period 2 myopic price.

3.5 Value of Returning Pricing

In this section, we develop a model to quantify the value of our proposed pricing model with returning customers, compared to a myopic pricing model which represents what is currently used in practice. Our main performance metric is the *revenue ratio*, which is defined as:

$$\text{revenue ratio} = \frac{\text{returning revenue}}{\text{myopic revenue}}.$$

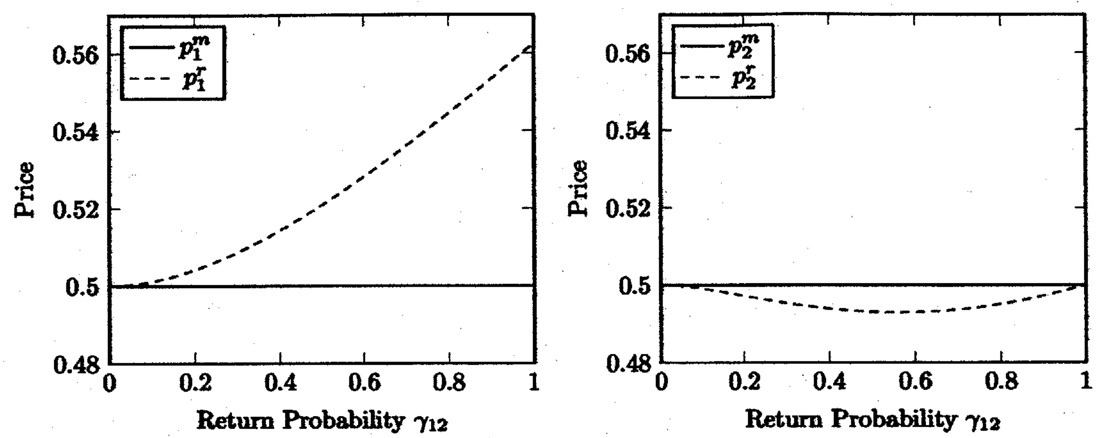


Figure 3-3: Myopic and Returning Prices from Example 3.4.8.

By definition, the revenue ratio is at least 1. If the revenue ratio were for example 1.02, this means that the returning customer pricing model gives a 2% revenue improvement compared to the myopic customer pricing model. The greater the revenue ratio, the larger the revenue improvement from using the returning customer pricing model.

We establish the assumptions which underlie our comparison in subsection 3.5.1, and formulate an optimization problem to compute the exact revenue ratio in subsection 3.5.2. In general, due to the constraints and the nonlinearity of the objective function, the optimization problem in subsection 3.5.2. is difficult and cannot be solved in closed form. Therefore, we develop bounds on the value of the returning pricing model which are functions of the returning probabilities and the first-time demand functions. In subsection 3.5.3, we develop upper bounds for the case when the first-time demand functions are linear; and in subsection 3.5.4, for the case when the first-time demand functions are nonlinear.

3.5.1 A Fair Comparison of the Two Models

We assume that the returning demand model is the true underlying demand model.

We wish to compare the revenue obtained by using the returning pricing model with the revenue obtained by using the myopic pricing model. The latter represents what is currently used in practice. The myopic demand model assumes that there are no returning customers, so that the demand in a given period depends only on the price in that period.

The value of our proposed returning pricing model is the incremental revenue obtained from the returning prices, relative to the revenue obtained from the myopic prices. Thus, the value of our model depends crucially on our baseline, which is the choice of the myopic demand model.

One clear possibility for the myopic demand model is to use the first-time demand functions as the myopic demand functions, i.e. $d_t^m(p_t) = f_t(p_t)$, and solving the optimization problem:

$$\begin{aligned} \max_{p_t} \quad & \sum_{t=1}^{T} p_t f_t(p_t) \\ \text{s.t} \quad & \sum_{t=1}^{T} f_t(p_t) \leq I \\ & p_1 \geq p_2 \geq \cdots \geq p_T \end{aligned}$$

This is equivalent as setting all of the return probabilities γ_{tu} to zero in (RMDO). However, this approach is problematic for the following reasons.

1. This approach assumes that the myopic demand functions are estimated using only first-time customers. However, in reality, in each period a retailer will observe the sales from both first-time customers and returning customers.

2. If the myopic demand functions are estimated using only the first-time demand, then the myopic demand functions are "wrong" in that the demand forecast at any period beyond the first period will not coincide with the actual realized demand. In fact, the myopic sales forecasts will be systematically lower than the true demand, which is composed of both first-time and returning demand.

In order to perform a fair comparison of the returning revenue and the myopic revenue, we instead make the following assumption that the myopic demand function is correct conditional on the past prices being the myopic prices.

Assumption 7 (Past myopic prices give myopic demand function). The myopic demand function in period t is equal to the returning demand function in period t given that the prices $p_1, \ldots, p_{t-1}$ are equal to the myopic prices $p_1^m, \ldots, p_{t-1}^m$, i.e.

$$d_t^m(p_t) = d_t(p_1^m, \ldots, p_{t-1}^m, p_t).$$

We illustrate Assumption 7 using an example.

Example 3.5.1. Suppose that the myopic demand functions are linear, i.e. $d_t^m(p_t) = a_t - b_t p_t$. Let p_1^m denote the myopic price in period 1 from solving the myopic pricing model. We will illustrate the consequences of Assumption 7 using Figure 3-4.

The demand in period 1 is only composed of first-time demand. At the price $p_1 = p_1^m$, the demand represented by the shaded triangle does not purchase in period 1 and could potentially return in future periods. If we assume for the sake of illustration that all of the customers who do not purchase in period 1 return in period 2, then Assumption 7 states that the total demand in period 2 is linear and a function only of p_2. This means that the first-time demand plus the returning demand is a linear function $d_2^m(p_2) = a_2 - b_2 p_2$.

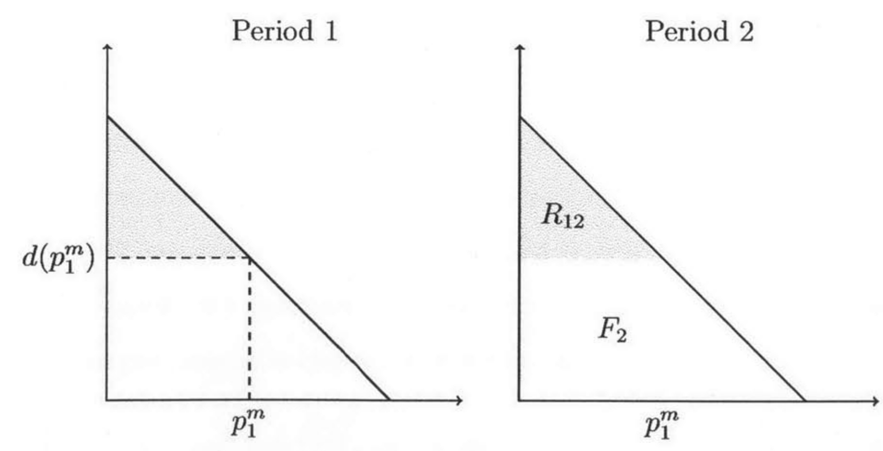


Figure 3-4: Illustration of Assumption 7

Note. The returning demand from period 1 in period 2 is represented by the triangle R_{12}, while the first-time demand in period 2 is represented by the white trapezoid F_2.

In addition to Assumption 7, in this section we make the simplifying assumption that the return probabilities are homogenous and can be characterized by a single parameter γ.

Assumption 8 (Homogeneous return probability). In the homogenous return probability demand model, the return probability γ_{tu} is given by

$$\gamma_{tu} = \begin{cases} \gamma & \text{if } u = t+1, \\ 0 & \text{otherwise.} \end{cases}$$

Assumption 8 states that a customer who does not purchase the product in period t either returns in period $t+1$ with probability γ, or does not return at any future period with probability $(1-\gamma)$. Conditional on the customer returning in period $t+1$, the customer either returns in period $t+2$ with the probability γ, or does not return at any future period with probability $(1-\gamma)$. This continues for all future periods.

Assumption 8 is convenient for us to characterize the value of returning pricing as a function of a single parameter.

3.5.2 Reformulation

As a consequence of Assumption 7 and Assumption 8, we have the following reformulation of the returning markdown optimization problem (RMDO).

Proposition 3.5.2. *Under Assumption 7 and Assumption 8, the returning markdown optimization problem* (RMDO) *can be reformulated as the following optimization problem:*

Input data:

- $I =$ *the initial inventory level.*
- $d_t^m =$ *the myopic demand function in period* t.

Decision variables:

- p_t = *the retail price of the item in period* t.
- s_t = *the sales in period* t.
- $r^+_{t,u}, r^-_{t,u}$ = *respectively the incremental and reduction in returning demand from customers who last visited the product in period* t *and first return to visit the product in period* u*, relative to the returning demand from these customer if the retail price in period* t *were set to the myopic price* p^m_t*, i.e.* $p_t = p^m_t$.

$$
\begin{aligned}
\max \quad & \sum_{t=1}^{T} p_t s_t && && \text{(3.17a)}\\
\text{s.t.} \quad & \sum_{t=1}^{T} s_t \leq I && && \text{(3.17b)}\\
& s_1 \leq d^m_1(p_1) && && \text{(3.17c)}\\
& s_t \leq d^m_t(p_t) - r^-_{t-1,t} + \sum_{u=1}^{t-1} r^+_{u,t} && \forall t = 2,3,\ldots,T && \text{(3.17d)}\\
& r^+_{t,u} \leq \gamma^{u-t}[d^m_t(p_u) - d^m_t(p_{u-1})] && \forall t \in \mathbb{T}, u = t+1, t+2, \ldots, T && \text{(3.17e)}\\
& -\frac{r^-_{t,t+1}}{\gamma} + \sum_{u=t+1}^{T} \frac{r^+_{t,u}}{\gamma^{u-t}} \leq d^m_t(p^m_t) - d^m_t(p_t) && \forall t \in \mathbb{T} && \text{(3.17f)}\\
& p_1 \geq p_2 \geq \cdots \geq p_T \geq 0 && && \text{(3.17g)}
\end{aligned}
$$

Discussion. The objective (3.17a) is to maximize the revenue. Constraint (3.17b) is the inventory constraint. Constraint (3.17c) sets the demand as the upper bound on sales in period 1, while constraint (3.17d) sets the demand as the upper bound on sales for periods $2, 3, \ldots, T$, with demand equal to the sum of the myopic demand plus the incremental returning customers and subtracting the reduced returning customers.

Constraints (3.17e), and (3.17f) together enforce the incremental or reduced returning demand from period t to u. In the case when $p_t > p_t^m$, there is incremental returning demand potential from period t to future periods. The incremental returning demand in period u, $r_{t,u}^+$, has an upper bound based on the retail prices in period $u-1$ and u as enforced in (3.17e), and also has an upper bound based on the incremental potential returning demand from period t that was already depleted in the intermediate periods, i.e. $r_{t,t+1}^+, \ldots, r_{t,u-1}^+$. In this case, due to the markdown pricing constraint, it would be optimal to set $r_{t,t+1}^- = 0$ as this would increase the upper bound on s_{t+1} and potentially increase the revenue in period $t+1$, In the case when $p_t < p_t^m$, the returning demand from period t to $t+1$ is reduced. Due to the markdown pricing constraint, and (3.17f), it is optimal to set $r_{t,u}^+ = 0$ and $r_{t,t+1}^-$ to the reduced returning demand.

Finally, constraint (3.17g) is the markdown pricing constraint.

Remark. Proposition 3.5.2 has the property that if the myopic demand functions d_t^m are linear, then the constraints are linear constraints, which also implies that the feasible region is convex.

Proof of Proposition 3.5.2. We first show that the returning markdown optimization problem (RMDO) can be reformulated as the following optimization problem:

$$
\begin{aligned}
\max \quad & \sum_{t=1}^{T} p_t s_t \\
\text{s.t.} \quad & \sum_{t=1}^{T} s_t \leq I \\
& s_t = d_t^m(p_t) - r_{t-1,t}^- + \sum_{u=1}^{t-1} r_{u,t}^+ \\
& r_{t,u}^+ = \gamma^{u-t}\left[d_t^m(\max\{p_u, p_t^m\}) - d_t^m(\max\{p_{u-1}, p_t^m\})\right] \\
& r_{t,t+1}^- = \gamma\left[d_t^m(\min\{p_t, p_t^m\}) - d_t^m(p_t^m)\right] \\
& p_1 \geq p_2 \geq \cdots \geq p_T
\end{aligned}
\tag{3.18}
$$

To prove this, we first note that the demand in period t is the sum of the demand given past myopic prices and the sum of the incremental and reduced returning demand due to the past prices being different from the myopic prices, i.e.

$$d_u(p_1, \ldots, p_u) = d_u^m(p_u) + \sum_{t=1}^{u-1} \left[r_{tu}^+ + r_{tu}^-\right].$$

The value of the terms r_{tu}^+ and r_{tu}^- depend on whether the posted price in period t exceeds the myopic price in period t, i.e. $p_t > p_t^m$; or the posted price in period t is less than the myopic price in period t, i.e. $p_t \leq p_t^m$.

In the first case (when $p_t > p_t^m$), then there is additional potential returning demand at the end of period t, relative to the myopic demand. This is illustrated in Figure 3-5. Depending on the future posted prices $p_{t+1}, \ldots, p_T$, some of this additional potential demand will returns and purchase in future periods $t+1, \ldots, T$.

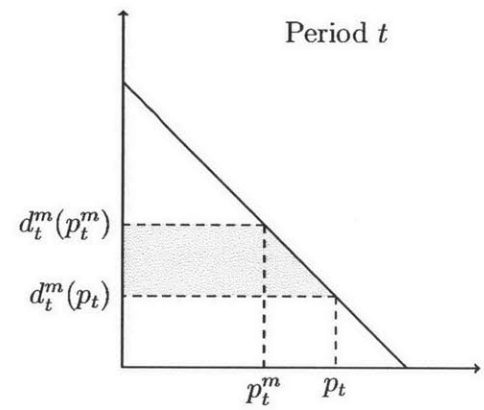


Figure 3-5: Illustration of Assumption 7 ($p_t > p_t^m$)

Note. If the price in period t, p_t, is greater than the myopic price in period t, p_{t+1}, then the additional potential returning demand is represented by the shaded trapezoid.

Recall that under the homogeneous return probability assumption, a proportion γ of demand that visits the product in period t returns in period $t+1$, while the remaining proportion $1-\gamma$ disappears forever.

The amount of returning demand realized in period $t+1$ depends on the ordering of the price in period $t+1$ relative to the myopic price in period t. If the price in period $t+1$, p_{t+1}, is less than the myopic price in period t, p_t^m, i.e. $p_{t+1} \leq p_t^m$, then the entire additional potential demand from period t that returns in period $t+1$ purchases the product in period $t+1$, i.e.

$$r^{+}_{t,t+1} = \gamma[d^m_t(p^m_t) - d^m_t(p_t)] \qquad \text{if } p_{t+1} \le p^m_t.$$

If the price in period $t+1$, p_{t+1}, is greater than the myopic price in period t, p^m_t, then only the customers with willingness to pay in the range $[p_{t+1}, p_t)$ will purchase in period $t+1$, i.e.

$$r^{+}_{t,t+1} = \gamma[d^m_t(p_{t+1}) - d^m_t(p_t)] \qquad \text{if } p_{t+1} > p^m_t.$$

We can combine these two cases in the expression

$$r^{+}_{t,t+1} = \gamma[d^m_t(\max\{p_{t+1}, p^m_t\}) - d^m_t(p_t)].$$

By applying the same logic, we can show more generally that for $u > t$,

$$r^{+}_{t,u} = \gamma^{u-t}[d^m_t(\max\{p_u, p^m_t\}) - d^m_t(\max\{p_t, p^m_t\})].$$

The above expression is correct even in the case when $p_t < p^m_t$, there is a reduction in returning demand relative to the myopic prices. In the case when $p_t < p^m_t$, both of the max terms evaluate to p^m_t and thus $r^{+}_{t,u}$ evaluates to zero.

In the second case (when $p_t > p^m_t$), there is a reduction in potential returning demand at the end of period t, relative to the myopic demand. This is illustrated in Figure 3-6.

Given that $p^m_t > p_t \ge p_{t+1}$, all of the returning demand would have purchased the product in period $t+1$. Therefore we have

$$r^{-}_{t,t+1} = \gamma[d^m_t(\min\{p_t, p^m_t\}) - d^m_t(p^m_t)] \le 0$$
$$r^{-}_{t,u} = 0 \qquad u = t+2, t+3, \ldots, T$$

In the case when $p_t \ge p^m_t$, there is no reduction in returning demand relative to myopic

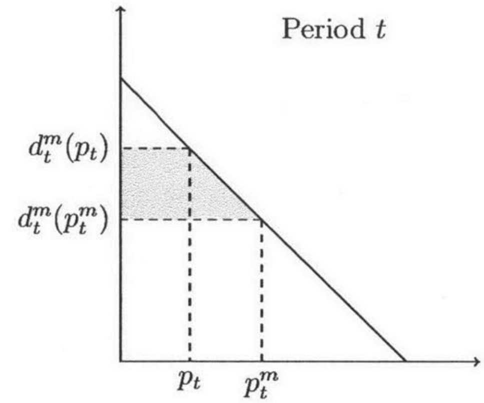


Figure 3-6: Illustration of Assumption 7 ($p_t < p_t^m$)

Note. If the price in period t, p_t, is less than the myopic price in period t, p_{t+1}, then the reduced potential returning demand is represented by the shaded trapezoid.

prices, and the expression

$$r_{t,u}^+ = \gamma^{u-t} \left[d_t^m(\max\{p_u, p_t^m\}) - d_t^m(\max\{p_{u-1}, p_t^m\})\right]$$

evaluates to zero.

Finally, it can be shown that (3.18) is equivalent to (3.17) based on the following observations:

1. The total incremental or reduction in the returning demand potential from period t to future periods u is $d_t^m(p_t^m) - d_t^m(p_t)$
2. In the case when $p_t > p_t^m$, because the prices are non-increasing, i.e. $p_t \geq p_{t+1} \geq \cdots \geq p_T$, the optimal solution of (3.17) sets $r_{t,u}^+$ to their correct value

$$r_{t,u}^+ = \gamma^{u-t} \left[d_t^m(\max\{p_u, p_t^m\}) - d_t^m(\max\{p_{u-1}, p_t^m\})\right].$$

3. In the case when $p_t < p_t^m$, because the prices are non-increasing, i.e. $p_t \geq p_{t+1} \geq \cdots \geq p_T$, the optimal solution of (3.17) sets $r_{t,u}^+$ to zero and

$$r_{t,t+1}^- = \gamma \left[d_t^m(p_t^m) - d_t^m(p_t)\right].$$

□

By solving the optimization problem (3.17), for given first-time demand functions d_t^m, we can obtain the revenue ratio for that particular setting. A plot of the revenue ratio for a class of linear demand functions with varying return probability is illustrated in Figure 3-7. We observe that the revenue ratio increases nonlinearly as the return probability increases, i.e. the increase in revenue by increasing the return probability from 40% to 50% is greater than the increase in revenue by increasing the return probability from 30% to 40%. This finding is a strong motivation for retailers to increase the return probability of customers (e.g. through marketing campaigns) as it has a significant impact on the revenue. Figure 3-7 also suggests that when the return probability is between 20–40% (which is what we observed from the data) then the returning customer pricing model yields a 1–3% increase in revenue relative to the myopic customer pricing model.

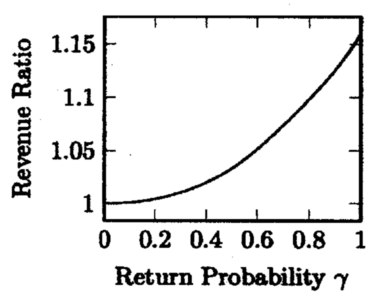


Figure 3-7: Revenue Ratio

Note. Demand parameters: $T = 3$ periods, linear first-time demand $d_t(p_t) = a_t - b_t p_t$, $a_1 = a_2 = a_3 = 100$, myopic prices $= a_t/2b_t = (50, 38, 26)$.

The previous section showed how to obtain the revenue ratio numerically. In the following two sections, we show how to derive insights about the value of returning pricing.

3.5.3 Linear Myopic Demand Functions

First-Time Demand in Period 1 Only

In the case when first-time customers only arrive in period 1, and the demand function in period 1 is linear, then we have a closed-form expression for the revenue ratio as stated in the following proposition.

Proposition 3.5.3. *In the case when there is only first-time demand in the first period, and the first-time demand function is linear, i.e.* $f_1(p) = a - bp$, *the value of the returning pricing model is given by*

$$\textit{revenue ratio} = \frac{H_t(\gamma)}{4\sum_{t=1}^{T}\left(\frac{\gamma}{4}\right)^{t-1}}.$$

To prove Proposition 3.5.3, we first note that the myopic prices are $p_t^m = a/2^t b$, which gives us

$$\text{myopic revenue} = \frac{a^2}{4b}\sum_{t=1}^{T}\left(\frac{\gamma}{4}\right)^{t-1}.$$

We also need the following result which determines the revenue of the optimal returning prices.

Proposition 3.5.4. *If first-time demand only arrives in period* 1, *and the first-time demand in period* 1 *is linear, i.e.* $f_1(p) = a - bp$, *then the optimal revenue for a* T*-period returning pricing problem* (3.17) *is* $H_T(\gamma)a^2/4b$ *where* $H_t(\gamma)$ *is defined by the following recurrence relation:*

$$H_t(\gamma) = \max_{0 \le p_t \le 1}\left\{p_t(1-p_t) + \gamma p_t^2 H_{t-1}(\gamma)\right\},$$

with $H_1(\gamma) = 1/4$.

In the proof of Proposition 3.5.4, it is convenient for us to define a normalized version of the problem, and scale the optimal solution for the normalized problem to an optimal solution for the regular problem.

Definition. A *T-period normalized returning pricing problem with linear first-time demand only in the first period* is defined as for a markdown pricing problem with T periods, where the first-time demand in period 1 is $f_1(p) = 1 - p$; and the first-time demand in periods $2, \ldots, T$ is zero; and customers are returning according to the homogeneous return probability assumption with parameter γ. It is convenient for us to number the time periods in reverse order, i.e. the first period is period T, the second period is period $T-1$, and so on. The mathematical formulation of the problem is:

$$\begin{aligned} J_T(\gamma) = \quad & \max \quad p_T(1-p_T) + \sum_{u=1}^{T-1} \gamma^{T-u} p_u(p_{u+1} - p_u) \\ & \text{s.t.} \quad 1 \geq p_T \geq p_{T-1} \geq \cdots \geq p_1 \geq 0 \end{aligned}$$

Proof of Proposition 3.5.4. We first show that $J_t(\gamma)$ satisfies the same recurrence relation as $H_t(\gamma)$ by following the chain of equalities:

$$\begin{aligned} & J_t(\gamma) \\ = & \max_{0 \leq p_1 \leq \cdots \leq p_t \leq 1} \Big\{ p_t(1-p_t) + \gamma p_{t-1}(p_t - p_{t-1}) \\ & \qquad + \gamma p_{t-2}(p_{t-1} - p_{t-2}) + \cdots + \gamma^{t-1} p_1(p_2 - p_1) \Big\} \\ = & \max_{0 \leq p_t \leq 1} \Big\{ p_t(1-p_t) + \gamma p_t^2 \max_{0 \leq p_1/p_t \leq \cdots \leq p_{t-1}/p_t \leq 1} \Big\{ \frac{p_{t-1}}{p_t}(1 - \frac{p_{t-1}}{p_t}) + \gamma \frac{p_{t-2}}{p_t}(\frac{p_{t-1}}{p_t} - \frac{p_{t-2}}{p_t}) \\ & \qquad + \cdots + \gamma^{t-2} \frac{p_1}{p_t}(\frac{p_2}{p_t} - \frac{p_1}{p_t}) \Big\} \Big\} \\ = & \max_{0 \leq p_t \leq 1} \Big\{ p_t(1-p_t) + \gamma p_t^2 J_{t-1}(\gamma) \Big\}. \\ = & \frac{1}{4(1 - \gamma J_{t-1}(\gamma))}. \end{aligned}$$

The above equalities show that $J_t(\gamma) = H_t(\gamma)$ and that by scaling the revenue of the normalized problem back to the original problem, the optimal revenue is $H_T(\gamma)a^2/4b$ as desired. $\square$

First-Time Demand in All Periods

We now consider the case when the first-time demand functions in the periods $t = 1, \ldots, T$ are linear, i.e. $f_t(p_t) = a_t - b_t p_t$. We derive an upper bound for the optimal value of the returning pricing model, by using the exact value we computed when the first-time demand only arrives in period 1.

We have the following upper bound on the optimal value of the returning pricing model.

Proposition 3.5.5. *If the myopic demand functions are linear, i.e. $d_t^m(p_t) = a_t - b_t p_t$, then an upper bound on the value of the returning pricing model is*

$$\textit{revenue ratio} \leq \frac{\sum_{t=1}^{T} H_{T-t+1}(\gamma) a_t^2/4b_t}{\sum_{t=1}^{T} a_t^2/4b_t} \leq H_T(\gamma).$$

Proof. We have an upper bound based on decoupling the objective function based on the first-time demand functions:

$$\begin{aligned} \max_{p_1 \geq \cdots \geq p_T} \left[\sum_{t=1}^{T} p_t f_t(p_t) + \sum_{t=1}^{T} \sum_{u=t+1}^{T} p_u \gamma^{u-t} [f_t(p_u) - f_t(p_{u-1})] \right] \leq \\ \sum_{t=1}^{T} \max_{p_t \geq \cdots \geq p_T} \left[p_t f_t(p_t) + \sum_{u=t+1}^{T} p_u \gamma^{u-t} [f_t(p_u) - f_t(p_{u-1})] \right] \end{aligned} \tag{3.19}$$

In the above inequality, the right hand side is the sum of T optimization problems, where the t-th optimization problem has the decision variables $p_t, p_{t+1}, \ldots, p_T$.

To derive the above inequality, one can repeatedly apply the inequality

$$\max_{x \in X} [f(x) + g(x)] \leq \max_{x \in X} f(x) + \max_{x \in X} g(x).$$

Next, we have by definition that

$$d_t^m(p_t) = f_t(p_t) + \sum_{u=1}^{t-1} \gamma_{ut} \left(d_u(p_1^m, \ldots, p_{u-1}^m, p_t) - d_u(p_1^m, \ldots, p_{u-1}^m, p_u^m) \right).$$

This implies that for every period t, the first-time demand function is less than the myopic demand function, i.e. $f_t \leq d_t^m$.

Consequently, we have the upper bound

$$
\begin{aligned}
&\max_{p_1 \geq \cdots \geq p_T} \left[\sum_{t=1}^{T} p_t f_t(p_t) + \sum_{t=1}^{T} \sum_{u=t+1}^{T} p_u \gamma^{u-t} [f_t(p_u) - f_t(p_{u-1})] \right] \\
\leq & \sum_{t=1}^{T} \max_{p_t \geq \cdots \geq p_T} \left[p_t d_t^m(p_t) + \sum_{u=t+1}^{T} p_u \gamma^{u-t} [d_t^m(p_u) - d_t^m(p_{u-1})] \right] \\
= & \sum_{t=1}^{T} \frac{a_t^2}{4b_t} H_{T-t+1}(\gamma).
\end{aligned}
$$

□

Intuitively, the proof of the proposition bounds the optimal value of the original problem by the sum of the optimal values of subproblems. The subproblem t can be interpreted as a $(T-t+1)$-period problem with linear first-time demand only in the first period. The upper bound is tight if first-time demand only arrives in the first period.

The upper bound defined in Proposition 3.5.5 is illustrated in Figure 3-8 and Figure 3-9. In these two examples, the maximum difference between the exact revenue ratio and the linear bound occurs at the return probability value $\gamma = 1$ and is approximately 19%. In our experiment based on real data (see section 3.6), we found that the return probability was in the range $[0.2, 0.4]$; and within this range the error is bounded below approximately 7%. The linear bound grows qualitatively at a similar order of magnitude as the exact revenue ratio.

3.5.4 Nonlinear Myopic Demand Functions

In order to define the upper bound for nonlinear myopic demand functions, we need to first introduce the two following assumptions.

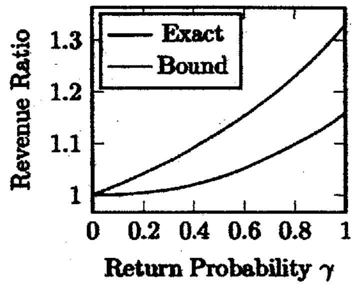


Figure 3-8: Revenue Ratio Bounds for Linear Demand (3 Periods)

Note. Demand parameters: $T = 3$ periods, linear first-time demand $d_t(p_t) = a_t - b_t p_t$, $a_1 = a_2 = a_3 = 100$, myopic prices $= a_t/2b_t = (50, 38, 26)$.

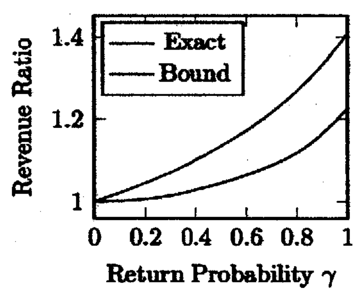


Figure 3-9: Revenue Ratio Bounds for Linear Demand (4 Periods)

Note. Demand parameters: $T = 4$ periods, linear first-time demand $d_t(p_t) = a_t - b_t p_t$, $a_1 = a_2 = a_3 = a_4 = 100$, myopic prices $= a_t/2b_t = (50, 46, 42, 38)$.

Assumption 9. For any time periods $u \leq t$, there exists an $\epsilon \geq 1$ such that

$$d_u^m(p) \leq \epsilon d_t^m(p) \qquad \forall p.$$

The above assumption means that even if from one period to another the myopic demand decreases for the same price, nevertheless it can be scaled up with $\epsilon \geq 1$.

Assumption 10. For any $u \leq t$, there exists $\delta \in (0,1]$ such that $d_u^m(p_{t-1}) \geq \delta d_u^m(p_t)$.

In fact, for any demand function, Lemma 3.5.6 below shows that there exists such a δ.

Lemma 3.5.6. *There exists a $\delta \in (0,1]$ such that $d_u^m(p_{t-1}) \geq \delta d_u^m(p_t)$.*

Proof. Let us define

$$\delta = \min_{t=1,\dots,T} \frac{d_u^m(p_{\max})}{d_u^m(p_{\min})}$$

Using the monotonicity of d_u^m, we note that $d_u^m(p_{t-1}) \leq d_u^m(p_{\min})$, $d_u^m(p_t) \geq d_u^m(p_{\max})$, this implies that

$$\frac{d_u^m(p_{t-1})}{d_u^m(p_t)} \geq \frac{d_u^m(p_{\max})}{d_u^m(p_{\min})} \geq \delta.$$

(To see that the first inequality is true, from the lhs to the rhs, we are decreasing the numerator and increasing the numerator.) □

Theorem 3.5.7. *Under Assumption 7 Assumption 8, Assumption 9 and Assumption 10, we have the upper bound*

$$\textit{revenue ratio} \leq 1 + \epsilon(1-\delta)\gamma(1+\gamma+\cdots+\gamma^{T-2}).$$

Discussion. When the demand functions decrease over time, (i.e. for $t < u$, $d_t(p)$ is a large multiplicative factor of $d_u(p)$), then ϵ is large, which makes the upper bound larger. As the minimum price increases to approch the maximum price, then δ increases to 1, which makes the upper bound smaller. The upper bound also increases as the return probability γ increases and as the number of time periods T increases. If $\gamma = 0$, the upper bound is 1 because the returning prices coincide with the myopic prices. If $\gamma = 1$, then the upper bound is $1 + \epsilon(1-\delta)\gamma T$.

Proof of Theorem 3.5.7. By applying the reformulation (3.18), we have

$$d_u(p_1, \ldots, p_u) = d_u^m(p_u) - r_{u-1,u}^- + \sum_{t=1}^{u-1} r_{tu}^+,$$

where $r_{t,u}^+, r_{u-1,u}^-$ denote the incremental and reduced returning demand due to the past prices being higher or lower than the myopic prices, and are defined by

$$\begin{aligned} r_{t,u}^+ &= \gamma^{u-t}\left[d_t^m(\max\{p_u, p_t^m\}) - d_t^m(\max\{p_{u-1}, p_t^m\})\right], \\ r_{t,t+1}^- &= \gamma\left[d_t^m(\min\{p_t, p_t^m\}) - d_t^m(p_t^m)\right]. \end{aligned}$$

$$\begin{aligned} d_t(p_1, \ldots, p_t) = d_t^m(p_t) &- \gamma\left[d_{t-1}^m(\max\{p_{t-1}, p_{t-1}^m\}) - d_{t-1}^m(\max\{p_{t-1}, p_{t-1}^m\})\right] \\ &+ \sum_{u=1}^{t-1} \gamma^{t-u}\left[d_u^m(\max\{p_t, p_u^m\}) - d_u^m(\max\{p_{t-1}, p_u^m\})\right] \end{aligned}$$

We next show that

$$d_u^m(\max\{p_t, p_u^m\}) - d_u^m(\max\{p_{t-1}, p_u^m\}) \leq \epsilon(1-\delta) d_u^m(p_t).$$

by considering the following cases.

Case 1: $p_u^m \geq p_{t-1} \geq p_t$ In this case, we get

$$d_u^m(\max\{p_t, p_u^m\}) - d_u^m(\max\{p_{t-1}, p_u^m\}) = d_u^m(p_u^m) - d_u^m(p_u^m) = 0.$$

Case 2: $p_{t-1} \geq p_u^m \geq p_t$ In this case, we get

$$d_u^m(\max\{p_t, p_u^m\}) - d_u^m(\max\{p_{t-1}, p_u^m\}) = d_u^m(p_u^m) - d_u^m(p_{t-1}).$$

Using the inequalities (i) $d_u^m(p_{t-1}) \geq \delta d_u^m(p_t)$ in Assumption 10 and (ii) $d_u^m(p_u^m) \leq d_u^m(p_t)$ in Assumption 9

$$d_u^m(p_u^m) - d_u^m(p_{t-1}) \leq d_u^m(p_t) - \delta d_u^m(p_t) = (1-\delta)d_u^m(p_t).$$

Case 3: $p_{t-1} \geq p_t \geq p_u^m$ In this case, we get

$$d_u^m(\max\{p_t, p_u^m\}) - d_u^m(\max\{p_{t-1}, p_u^m\}) = d_u^m(p_t) - d_u^m(p_{t-1}).$$

Again, using the inequality $d_u^m(p_{t-1}) \geq \delta d_u^m(p_t)$ in Assumption 10 gives us

$$d_u^m(p_t) - d_u^m(p_{t-1}) \leq d_u^m(p_t) - \delta d_u^m(p_t) = (1-\delta)d_u^m(p_t).$$

Therefore

$$\begin{aligned}
&\mathrm{Rev}(\mathbf{p}^r) \\
&= \sum_{t=1}^{T} p_t^r d_t^m(p_t^r) - \sum_{t=2}^{T} \gamma \left[d_{t-1}^m(\max\{p_{t-1}^r, p_{t-1}^m\}) - d_{t-1}^m(\max\{p_{t-1}^r, p_{t-1}^m\})\right] \\
&\quad + \sum_{t=2}^{T} p_t \sum_{u=1}^{t-1} \gamma^{t-u} \left[d_u^m(\max\{p_t^r, p_u^m\}) - d_u^m(\max\{p_{t-1}^r, p_u^m\})\right] \\
&\leq \sum_{t=1}^{T} p_t^r d_t^m(p_t) + \sum_{t=2}^{T} p_t^r \sum_{u=1}^{t-1} \gamma^{t-u} \epsilon(1-\delta) d_u^m(p_t^r) \\
&= \sum_{t=1}^{T} p_t^m d_t^m(p_t^m) + \sum_{t=2}^{T} \sum_{u=1}^{t-1} \gamma^{t-u} \epsilon(1-\delta) p_u^m d_u^m(p_u^m) \\
&= \sum_{t=1}^{T} p_t^m d_t^m(p_t^m) \left[1 + \epsilon(1-\delta) \sum_{u=t+1}^{T} \gamma^{t-u}\right] \\
&\leq \left[\sum_{t=1}^{T} p_t^m d_t^m(p_t^m)\right] \left[1 + \epsilon(1-\delta)(\gamma + \gamma^2 + \cdots + \gamma^{T-1})\right] \\
&= \mathrm{Rev}(\mathbf{p}^m) \left[1 + \epsilon(1-\delta)(\gamma + \gamma^2 + \cdots + \gamma^{T-1})\right]
\end{aligned}$$

□

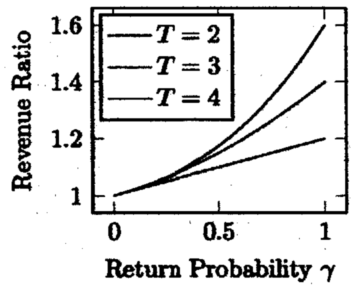


Figure 3-10: Revenue Ratio Bounds for Nonlinear Demand (Tighter Case)

Note. Demand parameters: $\epsilon = 1, \delta = 0.8$

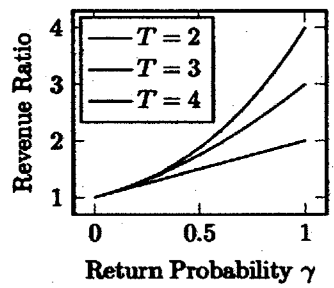


Figure 3-11: Revenue Ratio Bounds for Nonlinear Demand (Looser Case)

Note. Demand parameters: $\epsilon = 2, \delta = 0$

The nonlinear bounds are illustrated in Figure 3-10 and Figure 3-11. In Figure 3-10, the demand functions do not decrease with time, so that $\epsilon = 1$, and the maximum price is close to the minimum price, so that $\delta = 0.8$. In Figure 3-10, the demand functions do decrease with time, so that $\epsilon = 2$, and the maximum price is far from the minimum price, so that $\delta = 0$.

3.6 Model Estimation using Real Data

In this section, we analyze clickstream data obtained from an actual e-reatiler. We first seek to find evidence to support our hypothesis that a significant proportion of online shoppers are returning customers. We then perform a proof of concept experiment where we estimate a returning demand model using clickstream data, and whether the forecast accuracy is comparable to that of a myopic demand model estimated using aggregate data. Finally, we compare the prices and predicted revenues using prices from a returning demand model with prices from a myopic demand model in order to quantify the value of our model based on real data.

3.6.1 Validation with Data

Structure of Data

Clickstream data was provided to us by a customer analytics company. The data was obtained from an Internet fashion retailer. The clickstream data was a record of all activity by customers for five items in the sweaters category, over a time duration of three months (from 2012-09-01 to 2012-11-30). The data was a list of events, where each event has the following data fields:

- timestamp
- the customer cookie ID
- the item ID
- the event type
- the purchase price
- the number of units purchased

There were three types of customer events: viewing an item, putting the item in the shopping cart, and purchasing the item. The fields "purchase price" and "number of units purchased" are zero for non-purchase events.

Here are some interesting observations that we gleaned from the data:

- average proportion of the total sales that are due to returning customers 12.1%

- average proportion of the total revenue that is due to returning customers 19.4%
- average buy probability for a first-time customer 3.3%
- average buy probability for a returning customer 9.3%
- average probability that a first-time customer returns in a future period 5.0%

At first glance, because only 5.0% of customers who view an item return in a future period, it appears that the proportion of returning customers is small. However, many of the customers who view an item in an online store are just browsing with no intention of buying. If we were able to distinguish customers with some intention to purchase, we would find that a larger proportion than 5.0% of customers with intent to purchase return in a future period. Indeed, we find that customers who return to view an item are three times more likely to buy the item than customers who view the time for the first time. In addition, returning customers are responsible for 12.1% of sales and 19.4% of revenue. The reason why returning customers contribute more as a proportion of total revenue than number of units sold is because returning customers tend to be active earlier in the selling season when the price is higher.

Estimation Method

We estimated the myopic demand model

$$\log(s_{it}^{\text{total}}) = \sum_i \beta_j^1 \cdot ITEM_j(i) + \sum_t \beta_u^2 \cdot PERIOD_u(t) + \sum_i \beta_j^3 \cdot ELAST_j(i)\log(p_{it}) + \epsilon_{it},$$

where s_{it}^{total} is the sales of item i in period t, $ITEM_j(i)$, $PERIOD_u(t)$ and $ELAST_j(i)$ are indicator functions which take the value of 1 if and only if $j = i$ or $u = t$, and ϵ_{it} is assumed to be normally distributed i.i.d. random variables. The above demand model is a log-log demand model which is commonly used in practice (see e.g. [Talluri and van Ryzin, 2005]). This demand model assumes that the items share a common seasonality factor that depends on the period, and that each item has an item-specific intercept and item-specific elasticity.

The returning demand model forecasts the total sales for item i in period u is the sum of the first-time sales and the returning sales:

$$s_{it}^{\text{total}} = f_{it}(p_{it}) + \sum_{u<t} \lambda_{ut}[f_{iu}(p_{it}) - f_{iu}(p_{iu})]^{+}.$$

Step 1: As in the case of the myopic demand model, we estimate a log-log model for the first-time demand functions:

$$\log\left(\frac{s_{it}^{\text{first}}}{a_{it}^{\text{first}}}\right) \sim \sum_{j=1}^{n} \beta_j^1 ITEM_j(i) + \sum_{u=1}^{T} \beta_u^2 WEEK_u(t) + \sum_{j=1}^{n} \beta_j^3 ELAST_j(i) \log(p_{jt}) \quad (3.20)$$

Step 2: The observed estimated returning probability for item i is denoted by λ_{itu}, and is given by over items

$$\lambda_{itu} = \frac{s_{itu}^{\text{return}}}{[f_{it}(p_{iu}) - f_{it}(p_{it})]^{+}}.$$

We used the median of the set $\{\lambda_{itu}\}_i$ as the estimate for λ_{tu}.

For both the myopic and returning regression models, We used the `glmnet` package in the statistical programming language `R` to estimate a linear regression with constrained estimated elasticity values. This is often done in practice because if the estimated elasticities are large in absolute value, then the recommended prices tend to be the lowest permissible prices in the price ladder, which do not make sense in practice.

Correction factor. For both the myopic and returning demand models, the estimated log-log model tends to slightly underpredict sales, as measured by the sales bias or revenue bias (which are defined in section 3.6.1.) In order to correct the downward bias, we apply the correction factor described by Duan [1983]:

$$H^t = \frac{1}{|\mathbf{I}|} \sum_{i \in \mathbf{I}} \exp(\epsilon_{it})$$

where $\mathbf{I}$ is the set of item indices and ϵ_{it} are the error terms in the regression equation (3.20).

Forecast Metrics

We use the three forecast metrics which are commonly used in the literature. In order to define the metrics, let A_t denote the actual sales and F_t the forecasted sales.

- The volume-weighted mean absolute percentage error is defined by the formula

$$WMAPE = \frac{\sum_{t=1}^{T} |A_t - F_t|}{\sum_{t=1}^{T} A_t}.$$

  If the forecast is exact, i.e. $F_t = A_t$, then the WMPAE is equal to 0.

- The sales bias is defined by the formula

$$\text{sales bias} = \frac{\sum_{t=1}^{T} F_t}{\sum_{t=1}^{T} A_t}.$$

  If the forecast is exact, i.e. $F_t = A_t$, then the sales bias is equal to 1.

- The revenue bias is defined by the formula

$$\text{revenue bias} = \frac{\sum_{t=1}^{T} p_t F_t}{\sum_{t=1}^{T} p_t A_t}.$$

  If the forecast is exact, i.e. $F_t = A_t$, then the revenue bias is equal to 1.

Optimization Method

We assume that the estimated myopic and returning demand models are the true underlying demand model. For each item, we set the amount of inventory available equal to

110% of the total sales over the time horizon. We also imposed a markdown constraint over prices, i.e. $p_1 \geq p_2 \geq \cdots \geq p_6$.

We used a price ladder of $\{0.50, 0.55, \ldots, 1.00\}$. Recall that we had normalized the initial price of the item which was also the highest price to 1.00.

We compute the optimal markdown prices. We then compute the revenue gain for each item, which is defined as the percentage gain in the predicted revenues at the optimized prices to the predicted revenues at the actual prices.

3.6.2 Results and Discussion

The forecast metrics for the estimated myopic and returning demand models are shown in Table 3.1. The returning demand model has slightly better WMAPE than the myopic demand model. For the sales bias and revenue bias, the results are mixed. The mean and median sales and revenue bias for both demand models is fairly close to 100%. This suggests that the demand models do not have a strong upward or downward bias. The myopic demand model has a better mean bias, but the returning demand model has a better median bias. These results suggest that the returning demand model has forecast accuracy that is at least comparable to that of the myopic demand model.

The optimized revenue results are shown in Table 3.2. For the pricing results, if the myopic model were the true demand model, then the optimal myopic prices would result in a 4.8% increase in revenue over the actual prices; whereas if the returning model were the true demand model, then the optimal returning prices would result in a 6.3% increase in revenue over the actual prices. These results suggest that the returning demand model may generate more increase in revenue than the myopic demand model, by being able to extract more revenue from returning customers.

The mean optimized prices and the mean actual prices are shown in Figure 3-12. We note that averaged over the items, the actual price in period 1 was 1.00, while the actual price in period 6 was 0.70. The returning prices achieve approximately the same price dispersion as the actual prices, whereas the myopic prices are less dispersed.

Item	Myopic			Returning		
	WMAPE	Sales Bias	Revenue Bias	WMAPE	Sales Bias	Revenue Bias
1	0.1271	0.961	0.965	0.1774	1.008	1.013
2	0.1808	1.082	1.085	0.2109	1.067	1.067
3	0.1859	1.017	1.017	0.1608	0.920	0.925
4	0.2441	0.977	0.977	0.1817	1.003	1.000
5	0.1780	0.956	0.954	0.1785	0.910	0.911
Mean	0.1832	0.999	1.000	0.1819	0.982	0.983
Median	0.1808	0.977	0.977	0.1785	1.003	1.000
Std	0.0415	0.052	0.053	0.0182	0.066	0.065

Table 3.1: Forecast Metrics for Myopic and Returning Demand Models

Item	Myopic		Returning	
	$\frac{\text{rev}^{\text{myo}}(p^{\text{act}})}{\text{rev}^{\text{act}}}$	$\frac{\text{rev}^{\text{myo}}(p^{\text{myo}})}{\text{rev}^{\text{myo}}(p^{\text{act}})}$	$\frac{\text{rev}^{\text{ret}}(p^{\text{act}})}{\text{rev}^{\text{act}}}$	$\frac{\text{rev}^{\text{ret}}(p^{\text{ret}})}{\text{rev}^{\text{ret}}(p^{\text{act}})}$
1	96.56%	103.55%	101.67%	107.29%
2	108.44%	102.03%	107.43%	103.11%
3	101.45%	106.88%	93.22%	106.02%
4	97.69%	104.70%	100.58%	106.54%
5	95.40%	106.78%	91.40%	108.69%
Mean	99.91%	104.79%	98.86%	106.33%
Median	97.69%	104.70%	100.58%	106.54%

Table 3.2: Predicted Revenues for Myopic and Returning Demand Models.

Note. Note: the "actual prices" p^{act} are the actual prices rounded to the nearest price in the price ladder.

3.7 Conclusions

We investigate the impact of returning customers on pricing for fashion Internet retailers. Due to the convenience of online shopping, we expect that customers will visit online stores frequently and that customers might return to buy items at a lower price than the price which they saw the item listed at when they first viewed the item. By analyzing clickstream data from an online fashion retailer, we show that for this particular fashion retailer, price-sensitive returning customers—customers who first visit an

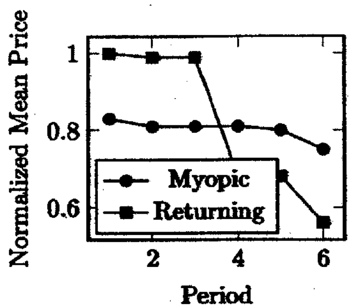


Figure 3-12: Mean myopic and returning prices per period.

item at a particular price, but purchase the item in a later visit at a lower price—are responsible for a significant proportion of sales and revenue.

We propose a demand model for fashion e-retailers which models returning customer behavior, can be estimated from clickstream data, which we incorporated into a pricing model. We also find conditions under which markdown prices are optimal for a dynamic pricing problem with returning customers, without imposing a markdown constraint.

One of the key insights of our work is that if a e-retailer is able to increase the proportion of returning customers, and incorporate returning customers into pricing decisions, the e-retailer can expect to see a substantial increase in revenue. We propose a model to make a fair comparison of the returning customer and myopic customer pricing models. Our model shows that when the return probability is in the range 20–40%, which is what we observed from the clickstream data, that the returning customer pricing model yields 1–3% more revenue than the myopic customer pricing model. We develop upper bounds on the value of the returning customer pricing model both for linear and nonlinear first-time demand functions. Finally, we perform numerical experiments based on real data, which suggest that the returning pricing model could achieve 2% more revenue compared to the myopic pricing model. This finding should motivate retailers to increase the returning customer probability, for example by using marketing campaigns, in order to increase their revenues and profits.

Bibliography

Hyun-soo Ahn, Mehmet Gümüs, and Philip Kaminsky. Pricing and manufacturing decisions when demand is a function of prices in multiple periods. *Operations Research*, 55(6):1039–1057, 2007.

Yossi Aviv and Amit Pazgal. Optimal pricing of seasonal products in the presence of forward-looking consumers. *Manufacturing & Service Operations Management*, 10 (3):339–359, 2008.

David Besanko and Wayne L Winston. Optimal price skimming by a monopolist facing rational consumers. *Management Science*, 36(5):555–567, 1990.

Gabriel Bitran and René Caldentey. An overview of pricing models for revenue management. *Manufacturing & Service Operations Management*, 5(3):203–229, 2003.

Gérard P Cachon and Robert Swinney. Purchasing, pricing, and quick response in the presence of strategic consumers. *Management Science*, 55(3):497–511, 2009.

Ronald H Coase. Durability and monopoly. *Journal of Law & Economics*, 15(1):143–149, 1972.

J Correa, Ricardo Montoya, and Charles Thraves. Contingent preannounced pricing policies with strategic consumers. Technical report, Working Paper, University of Chile, 2011.

Don Davis. E-commerce accounts for all the Q1 growth at Gap. `http://www.internetretailer.com/2014/05/22/e-commerce-accounts-all-q1-growth-gap`, 2014.

Naihua Duan. Smearing estimate: a nonparametric retransformation method. *Journal of the American Statistical Association*, 78(383):605–610, 1983.

Wedad Elmaghraby and Pınar Keskinocak. Dynamic pricing in the presence of inventory considerations: Research overview, current practices, and future directions. *Management Science*, 49(10):1287–1309, 2003.

Gadi Fibich, Arieh Gavious, and Oded Lowengart. Explicit solutions of optimization models and differential games with nonsmooth (asymmetric) reference-price effects. *Operations Research*, 51(5):721–734, 2003.

Scott C Friend and Patricia H Walker. Welcome to the new world of merchandising. *Harvard Business Review*, pages 133–142, 2001.

P Ghemawat and J L Nueno. Zara: Fast fashion. Harvard Business School Multimedia Case 9-703-416, 2003.

Aliza Heching, Guillermo Gallego, and Garrett van Ryzin. Mark-down pricing: An empirical analysis of policies and revenue potential at one apparel retailer. *Journal of Revenue and Pricing Management*, 1(2):139–160, 2002.

Daniel Kahneman and Amos Tversky. Prospect theory: An analysis of decision under risk. *Econometrica: Journal of the Econometric Society*, pages 263–291, 1979.

Praveen K Kopalle, Ambar G Rao, and Joao L Assuncao. Asymmetric reference price effects and dynamic pricing policies. *Marketing Science*, 15(1):60–85, 1996.

Qian Liu and Garrett J van Ryzin. Strategic capacity rationing to induce early purchases. *Management Science*, 54(6):1115–1131, 2008.

Javad Nasiry and Ioana Popescu. Dynamic pricing with loss-averse consumers and peak-end anchoring. *Operations research*, 59(6):1361–1368, 2011.

Charles Nicholls. The science of shopping cart abandonment. `http://seewhy.com/resources/shopping-cart-abandonment/`, 2014.

Nikolay Osadchiy and Gustavo Vulcano. Selling with binding reservations in the presence of strategic consumers. *Management Science*, 56(12):2173–2190, 2010.

Robert Phillips. *Pricing and revenue optimization*. Stanford University Press, 2005.

Ioana Popescu and Yaozhong Wu. Dynamic pricing strategies with reference effects. *Operations Research*, 55(3):413–429, 2007.

Stephen A Smith. Clearance pricing in retail chains. In *Retail Supply Chain Management*, pages 271–291. Springer, 2009.

Kalyan T Talluri and Garrett J van Ryzin. *The theory and practice of revenue management*, volume 68. Springer, 2005.

Chapter 4

Promotion Optimization for Grocery Retailers

4.1 Introduction

Sales promotions have became ubiquitous in various settings that include the grocery industry. During a sales promotion, the retail price of an item is temporarily lowered from the regular price, often leading to a dramatic increase in sales volume. To illustrate how important promotions are in the grocery industry, we consider a study conducted by A.C. Nielsen, which estimated that during January–June 2004, 12–25% of supermarket sales in five big European countries were made during promotion periods.

Our own analysis also supports the position that promotions are important and can be a key driver of increasing profits. We were able to obtain sales data from a large supermarket retailer for different categories of items. In Figure 4-1, we show the (normalized) prices and resulting sales for a particular brand of coffee in a single grocery store during a period of 35 weeks. One can see that this brand was promoted 8 out of 35 weeks (i.e., 23% of the time considered). In addition, the sales during promotions accounted for 41% of the total sales volume. Using a demand model estimated from real data (see Section 4.7.3 for details), we observe that the promotion prices of the retailer achieved a profit gain of 3% compared to using only the regular price (i.e., no promotions). A paper published by the Community Development Financial Institutions (CDFI) Fund reports that the average profit margin for the supermarket industry was 1.9% in 2010. According to analysis of Yahoo! Finance data, the average net profit margin for publicly traded US-based grocery stores for 2012 is close to 2010's 1.9% average. As a result, our finding suggests that promotions might make a significant difference in the retailer's profits. Furthermore, it motivates us to build a model that answers the following question: How much money does the retailer leave on the table by using the implemented prices relative to "optimal" promotional prices?

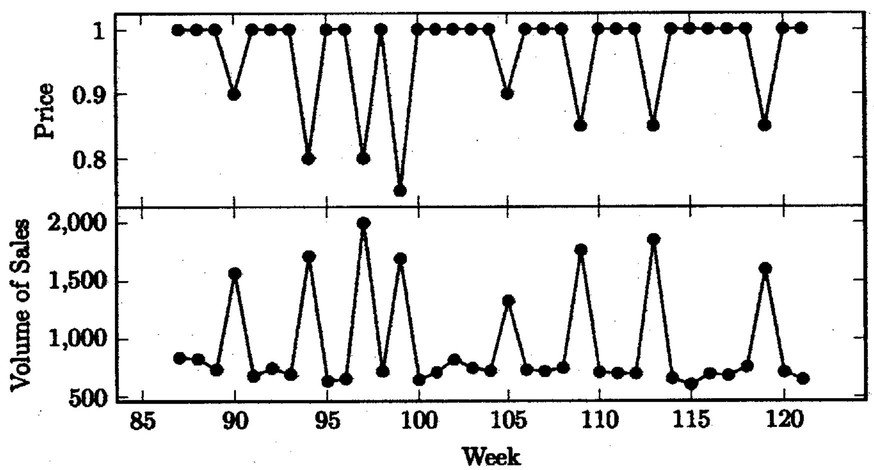


Figure 4-1: Prices and Sales of a Brand of Coffee

Given the importance of promotions in the grocery industry, it is not surprising that supermarkets pay great attention to how to design promotion schedules. The promotion planning process is complex and challenging for multiple reasons. First, demand exhibits a promotion fatigue effect, i.e., for certain categories of products, customers stockpile products during promotions, leading to reduced demand following the promotion. Second, promotions are constrained by a set of business rules specified by the supermarket and/or product manufacturers. Example of business rules include prices chosen from a discrete set, limited number of promotions and separating successive promotions (more details are provided in Section 4.3.1). Finally, the problem is difficult even for a single store because of its large scale - an average supermarket has of the order of 40,000 SKUs, and the number of items on promotion at any point of time is about 2,000 leading to a very large scale number of decisions that has to be made.

Despite the complexity of the promotion planning process, it is still to this day performed manually in most supermarket chains. This motivates us to design and study promotion optimization models that can make promotion planning more efficient (reducing man-hours) and at the same time more profitable (increasing profits and revenues) for supermarkets.

To accomplish this, we introduce a Promotion Optimization Problem (POP) formulation and propose how to solve it efficiently. We introduce and study classes of demand functions that incorporate the features we discussed above as well as constraints that model important business rules. The output will provide optimized prices together with performance guarantees. In addition, thanks to the scalability and the short running times of our formulation, the manager can test various what-if scenarios to understand the robustness of the solution.

The POP formulation we introduce is a nonlinear IP as a result, not computationally tractable, even for special instances. In practice, prices take values from a discrete price ladder (set of allowed prices at each time period) dictated by business rules. Even if we relax this requirement, the objective is in general neither concave nor convex due to the promotion fatigue effect. Since the objective of the POP is in general nonlinear, we propose a linear IP approximation and show that the problem can be solved efficiently as an LP. This new formulation approximates the POP problem for any general demand and hence, any desired objective function. We also establish analytical lower and upper bounds relative to the optimal objective that rely on the structure of the POP objective with respect to promotions. In particular, we show that when past prices have a multiplicative effect on current demand, for a certain subset of promotions, the profits are submodular in promotions, whereas when past prices have an additive effect, for all promotions the profits are supermodular in promotions. In other words, the results depend on the way that past prices affect demand rather than on the form of the demand function. These results allow us to derive guarantees on the performance of the LP approximation relative to the optimal POP objective. We also extend our analysis to the case of a combined demand model where both structures of past prices are simultaneously considered. Finally, we show using actual data that the models run fast in practice and can yield increased profits for the retailer by maintaining the same business rules.

The impact of our models can be also significant for supermarkets in practice. One of the goals of this research has been in fact to develop data driven optimization models that can guide the promotion planning process for grocery retailers, including the clients of Oracle Retail. They span the range of Mid-market (annual revenue below \$1 billion) as well as Tier 1 (annual revenue exceeding \$5 billion and/or 250+ stores) retailers all over the world. One key challenge for implementing our models into software that can be used by grocery retailers is the large-scale nature of this industry. For example, a typical Tier 1 retailer has roughly 1000 stores, with 200 categories each containing 50-600 items. An important criterion for our models to be adopted by grocery retailers in practice, is that the software solution needs to run in the order of a few seconds up to a minute. This is what has prompted us to reformulate our model as we discussed above as an LP.

Preliminary tests using actual supermarket data, suggest that our model can increase profits by 3% just by optimizing the promotion schedule and up to 5% by slightly increasing the number of promotions allowed. If we assume that implementing the promotions recommended by our models does not require additional fixed costs (this seems to be reasonable as we only vary prices), then a 3% increase in profits for a retailer with annual profits of \$100 million translates into a \$3 million increase. As we previously discussed, profit margins in this industry are thin and therefore 3% profit improvement is significative.

Contributions

This research was conducted in collaboration with our co-authors and industry practitioners from the Oracle Retail Science group, which is a business unit of Oracle Corporation. One of the end outcomes of this work is the development of sales promotion analytics that will be integrated into enterprise resource planning software for supermarket retailers.

- *We propose a POP formulation motivated by real-world retail environments.* We introduce a nonlinear IP formulation for the single item POP. Unfortunately, this model is in general not computationally tractable, even for special instances. An important requirement from our industry collaborators is that an executive of a medium-sized supermarket (100 stores, ~200 categories, ~100 items per category) can run the tool (whose backbone is the model and algorithms we are developing in this paper) and obtain a high quality solution in a few seconds. This motivates us to propose an LP approximation.

- *We propose an LP reformulation that allows us to solve the problem efficiently.* We first introduce a linear IP approximation of the POP. We then show that the constraint matrix is totally unimodular and therefore, our formulation is tractable. Consequently, one can use the LP approximation we introduce to obtain a provably near-optimal solution to the original nonlinear IP formulation.

- *We introduce general classes of demand functions that capture promotion fatigue effects.* An important feature of the application domain is the promotion fatigue effect observed. We propose general classes of demand functions in which past prices have a multiplicative or an additive effect on current demand. These classes are generalizations of some models currently found in the literature, provide some extra modeling flexibility and can be easily estimated from data. We also propose a unified demand model that combines the multiplicative and additive models and as a result, can capture several consumer segments.

- ***We develop bounds on performance guarantees for multiplicative and additive demand functions.*** We derive upper and lower guarantees on the quality of the LP approximation relative to the optimal (but intractable) POP solution and characterize the bounds as a function of the problem parameters. We show that for multiplicative demand, promotions have a submodular effect (for some relevant subsets of promotions). This leads to the LP approximation being an upper bound of the POP objective. For additive demand, we determine that promotions have a supermodular effect so that the LP approximation leads to a lower bound of the POP objective. Finally, we show the tightness of these bounds.

- ***We validate our results using actual data and demonstrate the added value of our model.*** Our industry partners provided us with a collection of sales data from multiple stores and various categories from their clients. We apply our analysis to a few selected categories. In particular, we looked into coffee, tea, chocolate and yogurt. We first estimate the various demand parameters and then quantify the value of our LP approximation relative to the optimal POP solution. After extensive numerical testing with the clients' data, we show that the approximation error is in practice even smaller than the analytical bounds we developed. Our model provides supermarket managers recommendations for promotion planning with running times in the order of seconds. As the model runs fast and can be implemented on a platform like Excel, it allows managers to test and compare various strategies easily. By comparing the predicted profit under the actual prices to the predicted profit under our LP optimized prices, we quantify the added value of our model.

4.2 Literature review

Our work is related to four streams of literature: optimization, marketing, dynamic pricing and retail operations. We formulate the promotion optimization problem for a single item as a nonlinear mixed integer program (NMIP). In order to give users flexibility in the choice of demand functions, our POP formulation imposes very mild assumptions on the demand functions. Due to the general classes of demand functions we consider, the objective function is typically non-concave. In general, NMIPs are difficult from a computational complexity standpoint. Under certain special structural conditions (e.g., see Hemmecke et al. [2010] and references therein), there exist polynomial time algorithms for solving NMIPs. However, many NMIPs do not satisfy these special conditions and are solved using techniques such as Branch and Bound, Outer-Approximation, Generalized Benders and Extended Cutting Plane methods [Grossmann, 2002].

In a special instance of the POP when demand is a linear function of current and past prices and when discrete prices are relaxed to be continuous, one can formulate the POP as a Cardinality-Constrained Quadratic Optimization (CCQO) problem. It has been shown in [Bienstock, 1996] that a quadratic optimization problem with a similar feasible region as the CCQO is NP-hard. Thus, tailored heuristics have been developed in order to solve the problem (see for example, Bertsimas and Shioda [2009] and Bienstock [1996]).

Our solution approach is based on linearizing the objective function by exploiting the discrete nature of the problem and then solving the POP as an LP. We note that due to the general nature of demand functions we consider, it is not possible to use linearization approaches such as in Sherali and Adams [1998] or Fletcher and Leyffer [1994]. We refer the reader to the books by Nemhauser and Wolsey [1988] and Bertsimas and Weismantel [2005] for integer programming reformulation techniques to potentially address the non-convexities. However, we observe that most of them are not directly applicable to our problem since the objective of interest is a time-dependent neither convex nor concave function.

As we show later in this paper, the POP for the two classes of demand functions we introduce is related to submodular and supermodular maximization. Maximizing an unconstrained supermodular function was shown to be a strongly polynomial time problem (see e.g., Schrijver [2000]). However, in our case, we have several constraints on the promotions and as a result, it is not guaranteed that one can solve the problem efficiently to optimality. In addition, most of the proposed methods to maximize supermodular functions are not easy to implement and are often not very practical in terms of running time. Indeed, our industry collaborators request solving the POP in at most few seconds and using an available platform like Excel. Unlike supermodular, maximization of submodular functions is generally NP-hard (see for example McCormick [2005]). Several common problems, such as max cut and the maximum coverage problem, can be cast as special cases of this general submodular maximization problem under suitable constraints. Typically, the approximation algorithms are based on either greedy methods or local search algorithms. The problem of maximizing an arbitrary non-monotone submodular function subject to no constraints admits a $1/2$ approximation algorithm (see for example, Buchbinder et al. [2012] and Feige et al. [2011]). In addition, the problem of maximizing a monotone submodular function subject to a cardinality constraint admits a $1-1/e$ approximation algorithm (e.g., Nemhauser et al. [1978]). In our case, we propose an LP approximation that does not request any monotonicity or other structure on the objective function. This LP approximation also provides guarantees relative to the optimal profits for two general classes of demand. Nevertheless, these bounds are parametric and not uniform. To compare them to the existing methods, we compute in Section 4.7 the values of these bounds on different demand functions estimated with actual data.

Sales promotions are an important area of research in the field of marketing (see Blattberg and Neslin [1990] and the references therein). However, the focus in the marketing community is on modeling and estimating dynamic sales models (typically econometric or choice models) that can be used to derive managerial insights [Cooper et al., 1999, Foekens et al., 1998]. For example, Foekens et al. [1998] study parametric econometrics models based on scanner data to examine the dynamic effects of sales promotions.

It is widely recognized in the marketing community that for certain products, promotions may have a *pantry-loading* or a *promotion fatigue* effect, i.e., consumers may buy additional units of a product during promotions for future consumption (stock piling behavior). This leads to a decrease in sales in the short term. In order to capture the promotion fatigue effect, many of the dynamic sales models that are used in the marketing literature have demand as a function of not just the current price, but also affected by past prices [Ailawadi et al., 2007, Mela et al., 1998, Heerde et al., 2000, Macé and Neslin, 2004]. The demand models used in our paper can be seen as a generalization of the demand models used in these papers.

Our work is also related to the field of dynamic pricing (see for example, Talluri and van Ryzin [2005] and the references therein). An alternative method to model the promotion fatigue effect is a reference price demand model, which posits that consumers have a reference price for the product based on their memory of the past prices (see e.g, Chen et al. [2013], Popescu and Wu [2007], Kopalle et al. [1996], Fibich et al. [2003]). When consumers purchase the product, they compare the posted price to their internal reference price and interpret a discount or surcharge as a gain or a loss. The demand models considered in our paper can be seen as a generalization of the reference price demand models as it includes several parameters to model the dependence of current demand in past prices. In Chen et al. [2013], the authors analyze a single product periodic review stochastic inventory model in which pricing and inventory decisions are made simultaneously and demand depends not only on the current price but also a memory-based reference. Popescu and Wu [2007], Kopalle et al. [1996], Fibich et al. [2003] all study dynamic pricing with a reference price effect by considering an infinite horizon setting without incorporating business rules. In our paper, we consider how to set prices while adhering to business rules which are important in practice.

Finally, our work is related to the field of retail operations and more specifically pricing problems under business rules. Subramanian and Sherali [2010] study a pricing problem for grocery retailers, where prices are subject to inter-item constraints. Due to the nonlinearity of the objective, they propose a linearization technique to solve the problem. Caro and Gallien [2012] study a markdown pricing problem for a fashion retailer. In this case, the prices are constrained to be non-increasing, and items in the same group are restricted to have the same prices over time.

The remainder of the paper is structured as follows. In Section 4.3, we describe the model and assumptions we impose as well as the business rules required for our problem. In Section 4.4, we formulate the Promotion Optimization Problem. In Section 4.5, we present an approximate formulation based on a linearization of the objective function, which gives rise to a linear IP. We show that the IP can in fact be solved as an LP. In Section 4.6, we consider multiplicative and additive demand models and show bounds on the LP approximation relative to the optimal POP solution. Section 4.7 presents computational results using real data. Finally, we present our conclusions in Section 4.8. Several of the proofs of the different propositions and theorems are relegated to the Appendix.

4.3 Model and Assumptions

In what follows, we consider the Promotion Optimization Problem for a single item. Note that solving this problem is important as one can use the single item model as a subroutine for the multiple product case. However, we believe this direction is beyond the scope of this paper. The manager's objective is to maximize the total profits during some finite time horizon, whereas the decision variables are for each time period, whether to promote a product and what price to set (i.e., the promotion depth). In our formulation, we also incorporate various important real-world business requirements that should be satisfied (a complete description is presented in Section 4.3.1). We first introduce some notation:

- T - Number of weeks in the horizon (e.g., one quarter composed of 13 weeks).
- L - Limitation on the number of times we are allowed to promote.
- S - Number of separating periods (restriction on the separation time between two successive promotions).

- $\mathcal{Q} = \{q^0 > q^1 > \cdots > q^k > \ldots > q^K\}$ - Price ladder, i.e., the discrete set of admissible prices.

- q^0 - Regular (non-promoted) price, which is the maximum price in the price ladder.

- q^K - Minimum price in the price ladder.

- c_t - Unit cost of the item at time t.

The decision variables are the prices set at each time period denoted by $p_t \in \mathcal{Q}$. Since we are considering a set of discrete prices only (motivated by the business requirement of a finite price ladder, see Section 4.3.1), one can rewrite the price p_t at time t as follows:

$$p_t = \sum_{k=0}^{K} q^k \gamma_t^k, \tag{4.1}$$

where γ_t^k is a binary variable that is equal to 1 if the price q^k is selected from the price ladder at time t and 0 otherwise. This way, the decision variables are now the set of binary variables γ_t^k; $\forall t = 1, \ldots, T$ and $\forall k = 0, \ldots, K$, for a total of $(K+1)T$ variables. In addition, we require the following constraint to ensure that exactly a single price is selected at each time t:

$$\sum_{k=0}^{K} \gamma_t^k = 1; \qquad \forall t. \tag{4.2}$$

Finally, we consider a general time-dependent demand function denoted by $d_t(\mathbf{p}_t)$ that explicitly depends on the current price and up to M past prices $p_t, p_{t-1}, \ldots, p_{t-M}$ as well as on demand seasonality and trend. We will consider specific demand forms later in the paper. $M \in \mathbb{N}_0$ denotes the memory parameter that represents the number of past prices that affect the demand at time t:

$$d_t(\mathbf{p}_t) = h_t(p_t, p_{t-1}, \ldots, p_{t-M}). \tag{4.3}$$

We next describe the various business rules we incorporate in our formulation.

4.3.1 Business Rules

1. *Promotion fatigue effect.* It is well known that when the price is reduced, consumers tend to purchase larger quantities. This can lead to a larger consumption for particular products but also can imply a stockpiling effect (see, e.g., Ailawadi et al. [2007] and Mela et al. [1998]). In other words, for particular items, customers will purchase larger quantities for future consumption (e.g., toiletries or non-perishable goods). Therefore, due to the consumer stockpiling behavior, a sales promotion for a product increases the demand at the current period but also reduces the demand in subsequent periods, with the demand slowly recovering over time to the nominal level, that is no promotion (see Figure 4-2). We propose to capture this effect by a demand model that explicitly depends on the current price p_t and on the past prices $p_{t-1}, p_{t-2}, \ldots, p_{t-M}$. In addition, our models allow to have the flexibility of assigning different weights to reflect how strongly a past price affects the current demand. The parameter M represents the memory of consumers with respect to past prices and varies depending on several features of the item. In practice, the parameter M can be estimated from data (see Section 4.7).

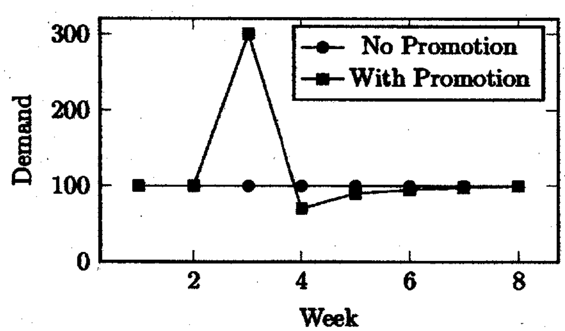


Figure 4-2: Illustration of the Promotion Fatigue Effect

Note. Promotion in week 3 yields a boost in current demand but also decreases demand in the following weeks. Finally, demand gradually recovers up to the nominal level (no promotion).

2. *Prices are chosen from a discrete price ladder.* For each product, there is a finite set of permissible prices. For example, prices may have to end with a '9'. In addition, the price ladder for an item can be time-dependent. This requirement is captured explicitly by equation (4.1), where the price ladder is given by: $q^0 > q^1 > \cdots > q^K$. In other words, the regular price q^0 is the maximal price and the price ladder has $K+1$ elements. For simplicity, we assume that the elements of the price ladder are time independent but note that this assumption can be relaxed.

3. *Limited number of promotions.* The supermarket may want to limit the frequency of the promotions for a product. This requirement applies because retailers wish to preserve the image of the store/brand. For example, it may be required to promote a particular product at most $L = 3$ times during the quarter. Mathematically, one can impose the following constraint in the formulation as follows:

$$\sum_{t=1}^{T}\sum_{k=1}^{K}\gamma_t^k \leq L. \tag{4.4}$$

4. *Separating periods between successive promotions.* A common additional requirement is to space out promotions by a minimal number of separating periods, denoted by S. Indeed, if successive promotions are too close to one another, this may hurt the store image and incentivize consumers to behave more as deal-seekers. Mathematically, one can impose the following constraint:

$$\sum_{\tau=t}^{t+S}\sum_{k=1}^{K}\gamma_\tau^k \leq 1 \qquad \forall t. \tag{4.5}$$

4.3.2 Assumptions

We assume that at each period t, the retailer orders the item from the supplier at a linear ordering cost that can vary over time, i.e., each unit sold in period t costs c_t. This assumption holds under the conventional wholesale price contract which is frequently used in practice as well as in the academic literature (see for example, Cachon and Lariviere [2005] and Porteus [1990]).

We also consider the demand to be specified by a deterministic function of current and past prices. This assumption is justified because we capture the most important factors that affect demand (current and past prices), therefore the estimated demand models are accurate in the sense of having low forecast error (see estimation results in Section 4.7 and Figure 4-7). Since the estimated deterministic demand functions seem to accurately model actual demand, for this application, we can use them as input into the optimization model without taking into account demand uncertainty.

Indeed, the typical process in practice is to estimate a demand model from data and then to compute the optimal prices based on the estimated demand model. In Section 4.7, we start with actual sales data from a supermarket, estimate a demand model and finally compute the optimal prices using our model. The demand models we consider are commonly used both by practitioners and the academic literature (see Heerde et al. [2000], Macé and Neslin [2004], Fibich et al. [2003]).

Finally, we assume that the retailer always carries enough inventory to meet demand, so that in each period, sales are equal to demand. The above assumption is reasonable in our setting because grocery retailers are aware of the negative effects of stocking out of promoted products (see e.g., Corsten and Gruen [2004] and Campo et al. [2000]) and use accurate demand estimation models (e.g., Cooper et al. [1999] and Van Donselaar et al. [2006]) in order to forecast demand and plan inventory accordingly. We hence use the terms demand and sales interchangeably in this paper.

To the best of our knowledge, this work is perhaps the first to develop a model that incorporates the aforementioned features for the POP and propose an efficient solution. These features not only introduce challenges from a theoretical perspective, but also are important in practice.

4.4 Problem Formulation

In what follows, we formulate the single-item Promotion Optimization Problem (POP) incorporating the business rules we discussed above:

$$
\begin{aligned}
\max_{\gamma_t^k} \quad & \sum_{t=1}^{T} (p_t - c_t) d_t(\mathbf{p_t}) \\
\text{s.t.} \quad & p_t = \sum_{k=0}^{K} q^k \gamma_t^k \\
& \sum_{t=1}^{T} \sum_{k=1}^{K} \gamma_t^k \leq L \\
& \sum_{\tau=t}^{t+S} \sum_{k=1}^{K} \gamma_\tau^k \leq 1 \qquad \forall t \\
& \sum_{k=0}^{K} \gamma_t^k = 1 \qquad \forall t \\
& \gamma_t^k \in \{0,1\} \qquad \forall t, k
\end{aligned}
\tag{POP}
$$

Note that the only decisions are which price to choose from the discrete price ladder at each time period (i.e., the binary variables γ_t^k). We denote by $POP(\mathbf{p})$ (or equivalently $POP(\gamma)$) the objective function of (POP) evaluated at the vector $\mathbf{p}$ (or equivalently γ). This formulation can be applied to a general time-dependent demand function $d_t(\mathbf{p_t})$ that explicitly depends on the current price p_t, and on the M past prices $p_{t-1}, \ldots, p_{t-M}$ as well as on demand seasonality and trend (see equation (4.3)). Specific examples are presented in Section 4.6.

The POP is a nonlinear IP (see Figure 4-3) and is in general hard to solve to optimality even for very special instances. Even getting a high-quality approximation may not be an easy task. First, even if we were able to relax the prices to take non-integer values, the objective is in general non-linear (neither concave nor convex) due to the cross time dependence between prices (see Figure 4-3). Second, even if the objective is linear, there is no guarantee that the problem can be solved efficiently using an LP solver because of the integer variables. We propose in the next section an approximation based on a linear programming reformulation of the POP.

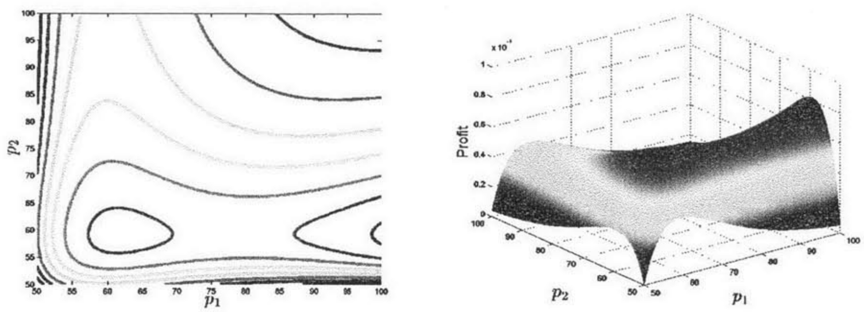


Figure 4-3: Profit Function for Demand with Promotion Fatigue Effect

Note. Parameters: Demand functions at time 1 and 2 follow the following relations: $\log d_1(p_1) = \log a_1 + \beta_1 \log p_1 + \beta_2 \log \frac{q^0 - p_1}{q^0}$; $\log d_2(p_2, p_1) = \log a_2 + \beta_1 \log p_2 + \beta_2 \log \frac{p_1 - p_2}{2q^0}$. Here, $a_1 = 100, a_2 = 200$ and $\beta_1 = -4$, $\beta_2 = 4$. The regular price, costs and minimum price are given by $q^0 = 100$, $c_1 = c_2 = 50$ and $q^K = 50$ respectively.

4.5 IP Approximation

By looking carefully at several data sets, we have seen that for many products, promotions often last only for one week, and two consecutive promotions are at least 3 weeks apart. If the promotions are subject to a separating constraint as in equation (4.5), then the interaction between successive promotions is fairly weak. Therefore, by ignoring the second-order interactions between promotions and capture only the direct effect of each promotion, we introduce a linear IP formulation that should give us a "good" solution. More specifically, we approximate the nonlinear POP objective by a linear approximation based on the sum of unilateral deviations. In order to derive the IP formulation of the POP, we first introduce some additional notation. For a given price vector $\mathbf{p} = (p_1, \ldots, p_T)$, we define the corresponding total profits throughout the horizon:

$$POP(\mathbf{p}) = \sum_{t=1}^{T} (p_t - c_t) d_t(\mathbf{p_t}).$$

Let us now define the price vector $\mathbf{p_t^k}$ as follows:

$$(\mathbf{p_t^k})_\tau = \begin{cases} q^k; & \text{if } \tau = t \\ q^0; & \text{otherwise} \end{cases}$$

In other words, the vector $\mathbf{p_t^k}$ has the promotion price q^k at time t and the regular price q^0 (no promotion) is used at all the remaining time periods. We also denote the regular price vector by $\mathbf{p^0} = (q^0, \ldots, q^0)$, for which the regular price is set at all the time periods. Let us define the coefficients b_t^k as:

$$b_t^k = POP(\mathbf{p_t^k}) - POP(\mathbf{p^0}). \tag{4.6}$$

These coefficients represent the unilateral deviations in total profits by applying a single promotion. One can compute these TK coefficients before starting the optimization procedure. Since these calculations can be done off-line, they do not affect the complexity of the optimization. We are now ready to formulate the IP approximation of the POP:

$$\begin{aligned} POP(\mathbf{p^0}) \quad & + \max_{\gamma_t^k} \sum_{t=1}^{T} \sum_{k=1}^{K} b_t^k \gamma_t^k \\ \text{s.t.} \quad & \sum_{t=1}^{T} \sum_{k=1}^{K} \gamma_t^k \leq L \\ & \sum_{\tau=t}^{t+S} \sum_{k=1}^{K} \gamma_\tau^k \leq 1 \qquad \forall t \\ & \sum_{k=0}^{K} \gamma_t^k = 1 \qquad \forall t \\ & \gamma_t^k \in \{0, 1\} \qquad \forall t, k \end{aligned} \tag{IP}$$

Remark (**Remark.**). One can condense the above IP formulation in a more compact way. In particular, since at most one of the decision variables $\{\gamma_t^k : k = 1, \ldots, K\}$ is equal to one, one can define $\tilde{b}_t = \max_{k=1,\ldots,K} b_t^k;\ \forall t = 1, \ldots, T$ and replace the double sums by single sums. As a result, we obtain a knapsack type formulation. Since both formulations are equivalent, we consider the (IP) above.

As we discussed, the IP approximation of the POP is obtained by linearizing the objective function. More specifically, we approximate the POP objective by the sum of the unilateral deviations by using a single promotion. Note that this approximation neglects the pairwise interactions of two promotions but still captures the promotion fatigue effect. We observe that the constraint set remains unchanged, so that the feasible region of both problems is the same. We also note that all the business rules from the constraint set are modeled as linear constraints. Consequently, the IP formulation is a linear problem with integer decision variables. As we mentioned, the IP approximation becomes more accurate when the number of separating periods S becomes large. In addition, the IP solution is optimal when there is no correlation between the time periods (i.e., when the demand at time t depends only on the current price and not on past prices) or when the number of promotions allowed is equal to one ($L = 1$). The instances where the IP is optimal are summarized in the following Proposition.

Proposition 4.5.1. *Under either of the following four conditions, the IP approximation coincides with the POP optimal solution. a) Only a single promotion is allowed, i.e.,* $L = 1$. *b) Demand at time t depends only on the current price* p_t *and not on past prices (i.e.,* $M = 0$*). c) The number of separating periods is at least equal to one (*$S \geq 1$*) and demand at time t depends on the current and last prices only (i.e.,* $M = 1$*). d) More generally, when the number of separating periods is at least the memory (i.e.,* $S \geq M$*).*

Proof. (a) When $L = 1$, only a single promotion is allowed and therefore the IP approximation is equivalent to the POP. Indeed, the IP approximation evaluates the POP objective through the sum of unilateral price changes.

(b) In the second case, demand at time t is assumed to depend only on the current price p_t and not on past prices. Consequently, the objective function is separable in terms of time (note that the periods are still tied together through some of the constraints). In this case too, the IP approximation is exact since each price change affects only the profit at the time it was made.

(c) We next show that the IP approximation is exact for the case where $S \geq 1$ and the demand at time t depends on the current and last period prices only.

Note that in this case, promotions affect only current and next period demands, but not demand in periods $t+2, t+3, \cdots$. We consider a price vector with two promotions at times t and u (i.e., $p_t = q^i$ and $p_u = q^j$) and no promotion at all the remaining times, denoted by $\mathbf{p}\{p_t = q^i, p_u = q^j\}$. From the feasibility with respect to the separating constraints, we know that t and u are separated by at least one time period. We need to show that the profits from doing both promotions is equal to the sum of the incremental profits from doing each promotion separately, that is:

$$\begin{aligned} &POP\big(\mathbf{p}\{p_t = q^i, p_u = q^j\}\big) - POP\big(\mathbf{p}^0\big) = \\ &\qquad POP\big(\mathbf{p}\{p_t = q^i\}\big) - POP\big(\mathbf{p}^0\big) + POP\big(\mathbf{p}\{p_u = q^j\}\big) - POP\big(\mathbf{p}^0\big). \end{aligned} \tag{4.7}$$

(d) One can extend the previous argument to generalize the proof for the case where the number of separating periods is larger or equal than the memory. Indeed, if $S \geq M$, the IP approximation is not neglecting correlations between different promotions and hence optimal. □

In general, solving an IP can be difficult from a computational complexity standpoint. In our numerical experiments, we observed that Gurobi solves (IP) in less than a second. The reason is that (IP) has an integral feasible region and therefore can be solved efficiently as an LP, as we show in the following Theorem. The feasible region of both (POP) and (IP) is given by:

$$\Big\{\gamma_t^k: \sum_{t=1}^{T}\sum_{k=1}^{K}\gamma_t^k \leq L \ \ \forall t; \ \sum_{\tau=t}^{t+S}\sum_{k=1}^{K}\gamma_t^k \leq 1; \ \sum_{k=0}^{K}\gamma_t^k = 1 \ \ \forall t\Big\}. \tag{4.8}$$

Theorem 4.5.2. *Every basic feasible solution of* (4.8) *is integral.*

Proof. We prove the result by expressing the LP relaxation of the IP in Linear Programming standard form, and then showing that the constraint matrix is totally unimodular.

We collect the decision variables γ_t^k, into a vector of size $(K+1)T$ as follows:

$$\gamma = [\gamma_1^0, \ldots, \gamma_1^K, \gamma_2^0, \ldots, \gamma_2^K, \ldots \gamma_T^0, \ldots, \gamma_T^K]^T.$$

Similarly, we denote by $\mathbf{b}$ the vectorization of the objective coefficients b_t^k defined in (4.6). By relaxing the integrality constraints, the IP problem can be written in the following standard LP form:

$$\begin{aligned} \max_{\gamma} \quad & \mathbf{b}^T\gamma \\ \text{s.t.} \quad & A\gamma \leq \mathbf{u} \\ & 0 \leq \gamma \leq 1 \end{aligned} \tag{4.9}$$

where $\mathbf{1}_K^T$ is a vector of ones with length K, and the matrix A and the vector $\mathbf{u}$ are given by:

$$A = \left[\begin{array}{l}
1 \;\; \mathbf{1}_K^T \\
\qquad\quad 1 \;\; \mathbf{1}_K^T \\
\qquad\qquad\qquad \ddots \\
\qquad\qquad\qquad\qquad\qquad \ddots \\
\qquad\qquad\qquad\qquad\qquad\qquad\quad \ddots \\
\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad \ddots \\
\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\quad 1 \;\; \mathbf{1}_K^T \\
\underbrace{0 \;\; \mathbf{1}_K^T \;\; 0 \;\; \mathbf{1}_K^T \;\; \cdots \;\; 0 \;\; \mathbf{1}_K^T \;\; 0 \;\; \mathbf{1}_K^T}_{S} \\
\qquad\quad \underbrace{0 \;\; \mathbf{1}_K^T \;\; \cdots \;\; 0 \;\; \mathbf{1}_K^T \;\; 0 \;\; \mathbf{1}_K^T \;\; 0 \;\; \mathbf{1}_K^T}_{S} \\
\qquad\qquad\qquad \ddots \qquad\qquad\qquad\qquad\qquad\quad \ddots \\
\qquad\qquad\qquad\qquad\quad \underbrace{0 \;\; \mathbf{1}_K^T \;\; 0 \;\; \mathbf{1}_K^T \;\; \cdots \qquad\quad 0 \;\; \mathbf{1}_K^T}_{S} \\
0 \;\; \mathbf{1}_K^T \;\; 0 \;\; \mathbf{1}_K^T \;\; \cdots \;\; \cdots \qquad\qquad\qquad \cdots \;\; \cdots \;\; 0 \;\; \mathbf{1}_K^T
\end{array}\right]; \quad \mathbf{u} = \begin{bmatrix} e_T \\ e_{T-S+1} \\ L \end{bmatrix}.$$

This matrix represents three different sets of constraints. The first T constraints are of the form $\sum_{k=0}^{K} \gamma_t^k = 1$ for each $t = 1, 2, \ldots, T$. We note that in (4.9), the equality is transformed to an inequality. This can be done because $b_t^0 = 0$ for all $t = 1, 2, \ldots, T$. Indeed, one can relax the equality in the initial integer formulation so that it allows the additional feasible solutions in which $p_t = 0$. Clearly, adding this new feasible solutions does not affect the optimality of the problem. The next set of $(T - S + 1)$ constraints represents the separating constraints from (4.5). Finally, the last row of A corresponds to the constraint on the limitation on the number of promotions allowed from (4.4).

To prove that matrix A is totally unimodular, we show that the determinant of any square sub-matrix B of A is such that $\det(B) \in \{-1, 0, +1\}$. Note that one can delete the columns corresponding to γ_t^0; $\forall t$ from the matrix A since these columns have only a single 1 entry. If we were to perform a Laplace expansion with respect to such a column, we would get the determinant of a smaller sub-matrix and therefore selecting those columns only multiplies the determinant by 1 or -1. After deleting these columns, we obtain a smaller matrix given by:

$$\tilde{A} = \begin{bmatrix}
\underbrace{\mathbf{1}_K^T \; \mathbf{1}_K^T \; \cdots \; \mathbf{1}_K^T \; \mathbf{1}_K^T}_{S} & & \\
& \underbrace{\mathbf{1}_K^T \; \cdots \; \mathbf{1}_K^T \; \mathbf{1}_K^T \; \mathbf{1}_K^T}_{S} & \\
& \ddots & \ddots \\
& & \underbrace{\mathbf{1}_K^T \; \mathbf{1}_K^T \; \cdots \; \mathbf{1}_K^T}_{S} \\
\mathbf{1}_K^T \; \mathbf{1}_K^T \; \cdots & & \mathbf{1}_K^T
\end{bmatrix}.$$

We observe that matrix $\tilde{A}$ has the consecutive-ones property. Therefore, matrix $\tilde{A}$ is totally unimodular and consequently every basic feasible solution of (4.8) is integral. □

Using Theorem 4.5.2, one can solve (IP) efficiently by solving its LP relaxation, given by:

$$\begin{aligned}
POP(\mathbf{p}^0) \quad & + \max_{\gamma_t^k} \sum_{t=1}^{T} \sum_{k=1}^{K} b_t^k \gamma_t^k \\
\text{s.t.} \qquad & \sum_{t=1}^{T} \sum_{k=1}^{K} \gamma_t^k \leq L \\
& \sum_{\tau=t}^{t+S} \sum_{k=1}^{K} \gamma_\tau^k \leq 1 \qquad \forall t \\
& \sum_{k=0}^{K} \gamma_t^k = 1 \qquad \forall t \\
& 0 \leq \gamma_t^k \leq 1 \qquad \forall t, k
\end{aligned} \tag{LP}$$

This allows us to obtain an approximation solution for the POP efficiently. From now on, we refer to (IP) as the LP approximation and denote its optimal solution by γ^{LP}. In addition, $LP(\mathbf{p})$ (or equivalently $LP(\gamma)$) denotes the objective function of (LP) evaluated at the vector $\mathbf{p}$ (or equivalently γ). The question is how does this LP approximation compare relative to the optimal POP solution. To address this question, we next consider two cases depending on the demand structure. First though, we propose some "reasonable" demand models in this application area.

4.6 Demand Models

In this section, we introduce two classes of demand functions. They incorporate the promotion fatigue effect we previously discussed. We next analyze supermarket sales data to support and validate the existence of the promotion fatigue effect in some items and categories. We report only a brief analysis here but a detailed description of the data will be presented in Section 4.7.

We divide the 117 weeks of data into a training set of 82 weeks and a testing set of 35 weeks. Below we consider a log-log demand model (see (4.32)). The latter is commonly used in industry (for example, by Oracle Retail) and in academia (see Heerde et al. [2000], Macé and Neslin [2004]). We then estimate two versions of the model. Model 1 is estimated under the assumption that there is no promotion fatigue effect, i.e., the memory parameter $M = 0$ in (4.32), so that the current demand d_t depends only on the current price p_t and not on past prices. Model 2 includes the promotion fatigue effect with a memory of two weeks, i.e., $M = 2$ in (4.32) so that the current demand d_t depends on the current price p_t and the prices in the two prior weeks p_{t-1} and p_{t-2}.

We summarize the regression results for a particular brand of coffee (the exact name of the brand cannot be explicitly unveiled due to confidentiality). We find that the estimated price elasticity coefficients of p_{t-1} and p_{t-2} for Model 2 are statistically significant. As a result, this supports the existence of the promotion fatigue effect for this item. In addition, we find that Model 2 has a significantly smaller forecast error relative to Model 1 (see Table 4.1). The estimated demand model for this coffee brand follows the following relation:

$$\log d_t = \beta^0 + \beta^1 t + \beta^2 WEEK_t - 3.277 \log p_t + 0.518 \log p_{t-1} + 0.465 \log p_{t-2}. \quad (4.10)$$

Here, β^0 and β^1 denote the brand intercept and the trend coefficient respectively. $\beta^2 = [\beta^2_t];\ t = 1, \ldots, 52$ is a vector with seasonality coefficients for each week of the year.

	Model 1	Model 2
MAPE	0.145	0.116
OOS R^2	0.827	0.900
Revenue Bias	1.069	1.059

Table 4.1: Forecast Metrics for Two Regression Models for a Brand of Coffee

Note. Model 1: No promotion fatigue effect. Model 2: Promotion fatigue with memory of 2 weeks. The forecast metrics MAPE, OOS R^2 and revenue bias are defined in Section 4.7.

In the remainder of this section, motivated by the above finding, i.e., that there are promotion fatigue effects in the demand, we introduce and study more general classes of demand models inspired by equation (4.10).

Notation

We introduce the following notation that will be used in the sequel. Let

$$A = \{(t_1, k_1), \ldots, (t_N, k_N)\}$$

with $N \leq L$ be a set of promotions with $1 \leq t_1 < t_2 < \cdots < t_N \leq T$. In other words, at each time period t_n; $\forall n = 1, \ldots, N$ the promotion price q^{k_n} is used, whereas at the remaining time periods, the regular price q^0 (no promotion) is set. It is convenient to define the price vector associated with the set A as:

$$(\mathbf{p}_A)_t = \begin{cases} q^{k_n} & \text{if } t = t_n \text{ for some } n = 1, \ldots, N; \\ q^0 & \text{otherwise.} \end{cases}$$

To further illustrate the above definition, consider the following example.

Example. Suppose that the price ladder is given by $Q = \{q^0 = 5 > q^1 = 4 > q^2 = 3\}$, and the time horizon is $T = 5$. Suppose that the set of promotions $A = \{(1,1), (3,2)\}$, that is we have two promotions at times 1 and 3 with prices q^1 and q^2 respectively. Then, $\mathbf{p}_A = (q^1, q^0, q^2, q^0, q^0) = (4, 5, 3, 5, 5)$. It is also convenient to define the indicator variables corresponding to the set of promotions A as follows:

$$(\gamma_A)_t^k = \begin{cases} 1 & \text{if } (\mathbf{p}_A)_t = q^k; \\ 0 & \text{otherwise.} \end{cases}$$

Note that matrix $(\gamma_A)_t^k$ has dimensions $(K+1) \times T$. In the previous example, we have:

$$\gamma_A = \begin{bmatrix} 0 & 1 & 0 & 1 & 1 \\ 1 & 0 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 \end{bmatrix},$$

Recall that the LP objective function is given by:

$$LP(\gamma) = POP(\mathbf{p}^0) + \sum_{t=1}^{T} \sum_{k=1}^{K} b_t^k \gamma_t^k, \tag{4.11}$$

where b_t^k is defined in (4.6). Finally, we denote by $\tilde{L}$ the effective maximal number of promotions given by:

$$\tilde{L} = \min\{L, \tilde{N}\}, \qquad \text{where } \tilde{N} = \left\lfloor \frac{T-1}{S+1} \right\rfloor + 1. \tag{4.12}$$

We assume that $L \geq 1$ (the case of $L = 0$ is not interesting as no promotions are allowed). Since $\tilde{N} \geq 1$, we also have $\tilde{L} \geq 1$.

4.6.1 Multiplicative Demand

In this section, we assume that past prices have a multiplicative effect on current demand, so that the demand at time t can be expressed by:

$$d_t = f_t(p_t) \cdot g_1(p_{t-1}) \cdot g_2(p_{t-2}) \cdots g_M(p_{t-M}). \tag{4.13}$$

Note that the current price elasticity along with the seasonality and trend effects are captured by the function $f_t(p_t)$. The function $g_k(p_{t-k})$ captures the effect of a promotion k periods before the current period, i.e., the effect of p_{t-k} on the demand at time t. M represents the memory of consumers with respect to past prices and can be estimated from data. As we verify in Section 4.7 from the actual data, it is reasonable to assume the following for the functions g_k.

Assumption 11. 1. Past promotions have a multiplicative reduction effect on current demand, i.e., $0 < g_k(p) \leq 1$.

2. Deeper promotions result in larger reduction in future demand, i.e., for $p \leq q$, we have: $g_k(p) \leq g_k(q) \leq g_k(q^0) = 1$.

3. The reduction effect is non-increasing with time after the promotion: g_k is non-decreasing with respect to k, i.e., $g_k(p) \leq g_{k+1}(p)$.

We assume that for $k > M$, $g_k(p) = 1 \ \forall p$, so that no effects are present after M periods.

Remark (**Remark.**). The demand in (4.13) represents a general class of demand models, which admits as special cases several models that are used in practice. For example, the demand model of Heerde et al. [2000] or Macé and Neslin [2004] with only pre-promotion effects that is of the form:

$$\log d_t = a_0 + a_1 \log p_t + \sum_{u=1}^{T} \log \beta_u \log p_{t-u}.$$

Next, we present upper and lower bounds on the performance guarantee of the LP approximation relative to the optimal POP solution for the demand model in (4.13).

Bounds on Quality of Approximation

Theorem 4.6.1. *Let γ^{POP} be an optimal solution to* (POP) *and let γ^{LP} be an optimal solution to* (LP). *Then:*

$$1 \leq \frac{POP(\gamma^{POP})}{POP(\gamma^{LP})} \leq \frac{1}{\underline{R}}, \tag{4.14}$$

where $\underline{R}$ is defined by:

$$\underline{R} = \prod_{i=1}^{\tilde{L}-1} g_{i(S+1)}(q^K), \tag{4.15}$$

with $\underline{R} = 1$ by convention, if $\tilde{L} = 1$.

Proof. Note that the lower bound follows directly from the feasibility of γ^{LP} for the POP. We next prove the upper bound by showing the following chain of inequalities:

$$\underline{R} \cdot LP(\gamma^{LP}) \overset{(i)}{\leq} POP(\gamma^{LP}) \overset{(ii)}{\leq} POP(\gamma^{POP}) \overset{(iii)}{\leq} LP(\gamma^{POP}) \overset{(iv)}{\leq} LP(\gamma^{LP}).$$

Inequality (i) follows from Proposition (4.6.3) below. Inequality (ii) follows from the optimality of γ^{POP} and inequality (iii) follows from part 2 of Lemma 4.6.2 below. Finally, inequality (iv) follows from the optimality of γ^{LP}. Therefore, we obtain:

$$\underline{R} = \underline{R} \cdot \frac{POP(\gamma^{POP})}{POP(\gamma^{POP})} \leq \underline{R} \cdot \frac{LP(\gamma^{LP})}{POP(\gamma^{POP})} \leq \frac{POP(\gamma^{LP})}{POP(\gamma^{POP})} \leq \frac{POP(\gamma^{POP})}{POP(\gamma^{POP})} = 1.$$

□

Theorem 4.6.1 relies on the following two results.

Lemma 4.6.2 (Submodular effect of the last promotion on profits).

1. *Let $A = \{(t_1, k_1), \ldots, (t_n, k_n)\}$ be a set of promotions with $t_1 < t_2 < \cdots < t_n$ $(n \leq L)$ and let $B \subset A$. Consider a new promotion (t', k') with $t_n < t'$. If the new promotion (t', k'), when added to A, yields larger profits than $\mathbf{p}_A$, that is:*

$$POP(\gamma_{A\cup\{(t',k')\}}) \geq POP(\gamma_A), \tag{4.16}$$

then the promotion (t', k') yields a larger marginal profit increase for $\mathbf{p}_B$ than for $\mathbf{p}_A$, that is:

$$POP(\gamma_{A\cup\{(t',k')\}}) - POP(\gamma_A) \leq POP(\gamma_{B\cup\{(t',k')\}}) - POP(\gamma_B). \tag{4.17}$$

2. *Let γ^{POP} be an optimal solution for the POP. Then: $POP(\gamma^{POP}) \leq LP(\gamma^{POP})$.*

Note that if (4.16) is not satisfied, the sub-additivity property of Lemma 4.6.2 does not necessarily hold for any feasible solution. However, the required condition in (4.16) is always automatically satisfied for the optimal POP solution. The proof of Lemma 4.6.2 can be found in Appendix 4.A. Lemma 4.6.2 states that for a multiplicative demand model as in (4.13), the POP profits are submodular in promotions (for certain relevant sets of promotions). Consequently, it supports intuitively the fact that the LP approximation overestimates the POP objective, i.e., $POP(\gamma^{POP}) \leq LP(\gamma^{POP})$.

Proposition 4.6.3. *For any feasible vector γ, we have: $POP(\gamma) \geq \underline{R} \cdot LP(\gamma)$.*

The proof of Proposition 4.6.3 can be found in Appendix 4.B. It provides a lower bound for the POP objective function by applying the linearization and compensating by the worst case aggregate factor, that is $\underline{R}$.

Using the results of Theorem 4.6.1, one can solve the LP approximation (efficiently) and obtain a guarantees relative to the optimal POP solution. These bounds are parametric and can be applied to any general demand model in the form of equation (4.13). In addition, as we illustrate in Section 4.6.1, these bounds perform well in practice for a wide range of parameters.

We next show that the bounds of Theorem 4.6.1 are tight.

Proposition 4.6.4 (Tightness of the bounds for multiplicative demand).

1. *The lower bound in Theorem 4.6.1 is tight. More precisely, for any given price ladder, L, S and functions g_k, there exist T, costs c_t and functions f_t such that:*

$$POP(\gamma^{POP}) = POP(\gamma^{LP}).$$

2. *The upper bound in Theorem 4.6.1 is asymptotically tight. For any given price ladder, S and functions g_k, there exists a sequence of promotion optimization problems $\langle POP^n \rangle_{n=1}^{\infty}$, each with a corresponding LP solution γ_n^{LP} and optimal POP solution γ_n^{POP} such that:*

$$\lim_{n \to \infty} \frac{POP^n(\gamma_n^{POP})}{POP^n(\gamma_n^{LP})} = \frac{1}{\underline{R}_{\infty}}.$$

The proof of Proposition 4.6.4 can be found in Appendix 4.C.

Illustrating the bounds

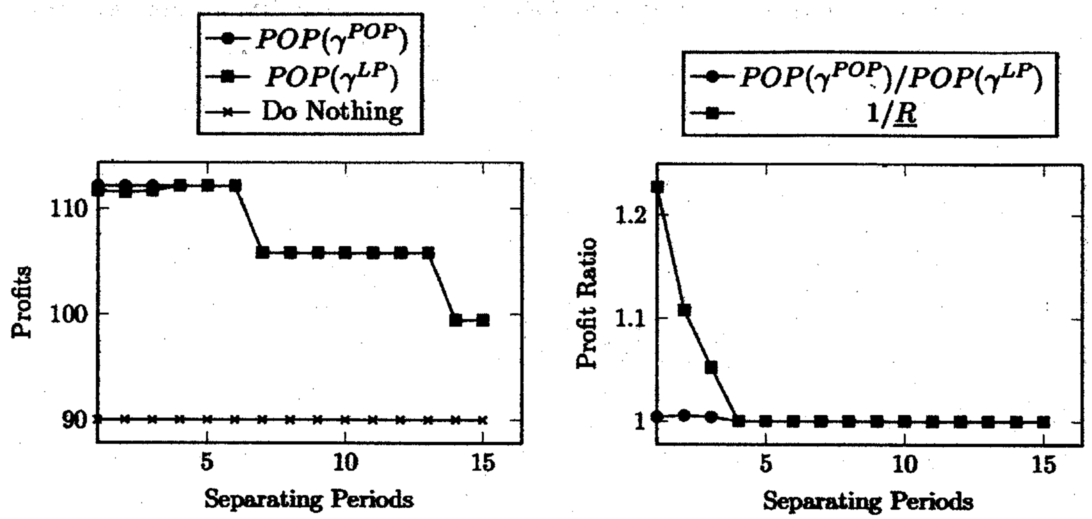
We show some examples that illustrate the behavior and quality of the bounds we have developed in the previous section. Recall that solving the POP can be hard in practice. Therefore, one can instead implement the LP solution. The resulting profit is then equal to $POP(\gamma^{LP})$, whereas in theory, we could have obtained a maximum profit equal to the optimal POP profits denoted by $POP(\gamma^{POP})$. In our numerical experiments, we examine the gap between $POP(\gamma^{LP})$ and $POP(\gamma^{POP})$ as a function of various parameters of the problem. In addition, we compare the ratio between $POP(\gamma^{POP})$ and $POP(\gamma^{LP})$ relative to the lower bound in Theorem 4.6.1 equal to $1/\underline{R}$. We also present an additional curve labeled "Do Nothing" as a benchmark (for which the no-promotion price is used at each time).

As we previously noted, the bounds we developed depend on four different parameters: the number of separating periods S, the number of promotions allowed L, the value of the minimum element of the price ladder q^K and the effect of past prices (i.e., the value of the memory parameter M as well as the magnitude of the functions g_k) . Below, we study the effect of each of these factors by varying them one at a time while the others are set to their worst case value.

All the figures below lead us to the following two observations: *a*) The LP solution achieves a profit that is close to the optimal profit. *b*) In particular, the actual optimality gap (between the POP objective at optimality versus evaluated at the LP approximation solution) seems to be of the order of 1-2 % and is smaller than the upper bound which we developed in Theorem 4.6.1.

In Figure 4-4 and Figure 4-5, the demand model we use is given by: $\log d_t(\mathbf{p}) = \log(10) - 4\log p_t + 0.5\log p_{t-1} + 0.3\log p_{t-2} + 0.2\log p_{t-3} + 0.1\log p_{t-4}$.

Dependence on separating periods: In Figure 4-4, we vary the number of separating periods S from 1 to 16 (remember that the horizon is $T = 35$ weeks). We make the following observations: *a*) As one would expect from Proposition 4.5.1, the LP approximation coincides with the optimal POP solution when $S \geq M = 4$, i.e., $S \geq 4$. *b*) Our intuition suggests that as S increases, the upper bound $1/\underline{R}$ becomes better. Indeed, the promotions are further apart in time, reducing the interaction between promotions and improving the quality of the LP approximation. *c*) For values of $S \geq 1$, the upper



(a) Profits

(b) Profit ratio

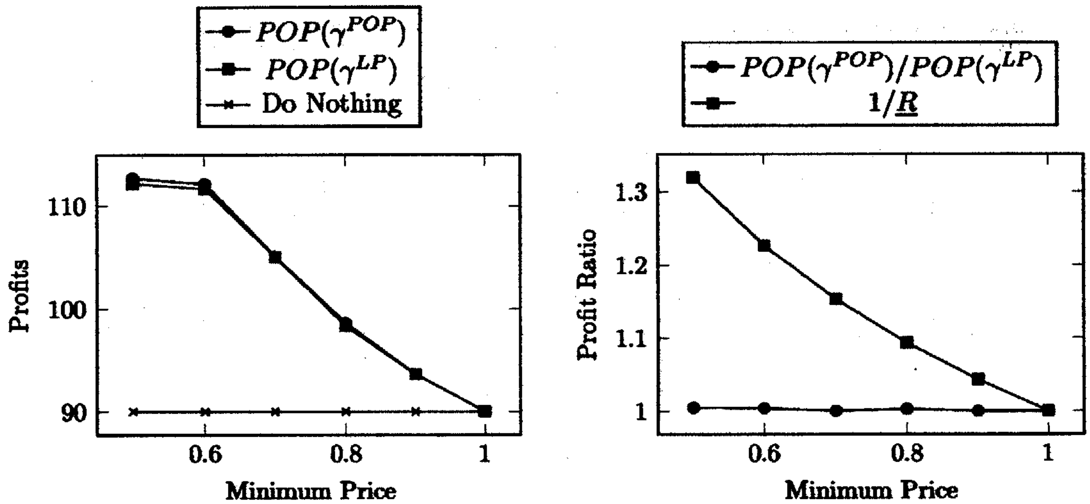
Figure 4-4: Results of Multiplicative Demand Model (Varying Separation)

Note. Example parameters: $L = 3, Q = \{1, 0.9, 0.8, 0.7, 0.6\}$.

bound is at most 23% in this example. In practice, typically the number of separating periods is at least 1 but often 2-4 weeks.

Dependence on the number of promotions allowed: In Figure ??, we vary the number of promotions allowed L between 0 and 8. We make the following observations: *a*) As one would expect from Proposition 4.5.1, the LP approximation coincides with the optimal POP solution when $L = 1$ (and of course $L = 0$). *b*) The upper bound is at most 23% in this example. Note that from the definition of $\underline{R}$ in equation (4.35) of Theorem 4.6.1, $1/\underline{R}$ increases with L up to $L = 3$. Indeed, since $S = 1$ and $M = 4$, the first promotion can never interact with the fourth promotion or with further ones.

Dependence on the minimal element of the price ladder: In Figure 4-5, we vary the (normalized) minimum promotion price q^K between 0.5 and 1. We make the following observations: *a*) As one would expect the LP approximation coincides with the optimal POP solution when $q^K = 1$, i.e., the promotion price is equal to the regular price so that promotions do not exist. *b*) The upper bound is 33% in this example



(a) Profits

(b) Profit ratio

Figure 4-5: Results of Multiplicative Demand Model (Varying Minimum Price)

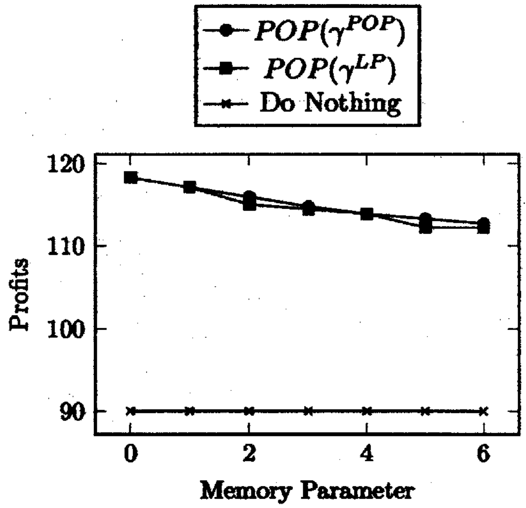
Note. Example parameters: $L = 3, S = 1$.

for the case where a 50% promotion is allowed. If we restrict to a maximum of 30% promotion price, the bound becomes 14%. Using the definition of $\underline{R}$ from (4.35), $1/\underline{R}$ decreases with q^K.

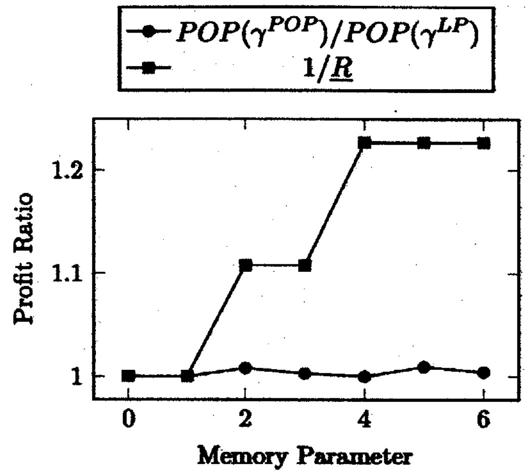
Dependence on the length of the memory: In Figure 4-6, we vary the memory of costumers with respect to past prices, M between 0 and 6. Note that in this example, we have chosen the functions $g_1, g_2, \ldots, g_M$ to be equal. This choice can be seen as the "worst case" so that past prices have a uniformly strong effect on current demand. We make the following observations: *a*) As one would expect from Proposition 4.5.1, the LP approximation coincides with the optimal POP solution when $S \geq M$, i.e., $M \leq 1$. *b*) The upper bound is 23% in this example. Using the definition of $\underline{R}$ from (4.35), $1/\underline{R}$ increases with M.

4.6.2 Additive Demand

Our analysis of the sales data suggests that for some products, one needs to consider a demand model where the effect of past prices on current demand is additive. Motivated



(a) Profits



(b) Profit ratio

Figure 4-6: Results of Multiplicative Demand Model (Varying Memory)

Note. Example parameters: $\log d_t(\mathrm{p}) = \log(10) - 4\log p_t + 0.2\log p_{t-1} + 0.2\log p_{t-2} + \cdots + 0.2\log p_{t-M}$; $L = 3, S = 1$.

by this observation, we also propose and study a class of additive demand functions. Suppose that past prices have an additive effect on current demand, so that the demand at time t is given by:

$$d_t = f_t(p_t) + g_1(p_{t-1}) + g_2(p_{t-2}) + \cdots + g_M(p_{t-M}). \tag{4.18}$$

As we verify in Section 4.7 from the actual data, it is reasonable to assume the following structure for the functions g_k.

Assumption 12. 1. The reduction effect is non-positive, i.e., $g_k(p) \leq 0$.

2. Deeper promotions result in larger reduction in future demand, i.e., $p \leq q$ implies that $g_k(p) \leq g_k(q) \leq g_k(q^0) = 0$.

3. The reduction effect is non-increasing with time since after the promotion: g_k is non-decreasing with respect to k, i.e., $g_k(p) \leq g_{k+1}(p)$.

Note that the above assumptions are analogous to Assumption 11 for the multiplicative model. We assume that for $k > M$, $g_k(p) = 0 \; \forall p$.

Remark (**Remark.**). Equation (4.18) represents a general class of demand functions, which admits as special cases several demand models used in practice. For example, the demand model used by Fibich et al. [2003] with symmetric reference price effects is given by:

$$d_t = a - \delta p_t - \phi(p_t - r_t). \tag{4.19}$$

Equation (4.19) can be rewritten as: $d_t = a - (\delta + \phi)p_t + \phi r_t$. Here, r_t represents the reference price at time t that consumers are forming based on their memory of past prices. The parameter ϕ denotes the price sensitivity with respect to the reference price, whereas $\delta + \phi$ represents the price sensitivity with respect to the current price. Note that the reference price at time t is given by:

$$r_t = (1 - \theta)p_{t-1} + \theta r_{t-1},$$

and can be rewritten in terms of past prices as follows:

$$r_t = (1 - \theta)p_{t-1} + \theta(1 - \theta)p_{t-2} + \theta^2(1 - \theta)p_{t-3} + \cdots = (1 - \theta)\sum_{k=1}^{T} \theta^{k-1} p_{t-k},$$

where $0 \leq \theta < 1$ denotes the memory of the consumers towards past prices. Therefore, the current demand from equation (4.19) can be written as follows in terms of the current and past prices:

$$d_t = a - (\delta + \phi)p_t + \sum_{k=1}^{M=T} (1 - \theta)\phi\, \theta^{k-1} p_{t-k}. \tag{4.20}$$

One can see that equation (4.20) falls under the model we proposed in (4.18), when the functions g_k are chosen appropriately and the memory parameter M goes to infinity. In addition, the additive model from (4.18) provides more flexibility in choosing the suitable memory parameter using data and allows us to give different weights depending on how far is the past promotion from the current time period.

Next, we present upper and lower bounds on the performance guarantee of the LP approximation relative to the optimal POP solution for the demand model in (4.18).

Bounds on Quality of Approximation

Theorem 4.6.5. *Let γ^{POP} be an optimal solution to* (POP) *and let γ^{LP} be an optimal solution to the LP approximation. Then:*

$$1 \leq \frac{POP(\gamma^{POP})}{POP(\gamma^{LP})} \leq 1 + \frac{\overline{R}}{POP(\gamma^{LP})}. \tag{4.21}$$

where $\overline{R}$ is defined by:

$$\overline{R} = \sum_{i=1}^{\tilde{L}} \sum_{j=i+1}^{\tilde{L}} (q^K - q^0) g_{(j-i)(S+1)}(q^K). \tag{4.22}$$

Proof. Note that the lower bound follows directly from the feasibility of γ^{LP} to the POP. We next prove the upper bound by showing the following chain of inequalities:

$$LP(\gamma^{LP}) \overset{(i)}{\leq} POP(\gamma^{LP}) \overset{(ii)}{\leq} POP(\gamma^{POP}) \overset{(iii)}{\leq} LP(\gamma^{POP}) + \overline{R} \overset{(iv)}{\leq} LP(\gamma^{LP}) + \overline{R}. \tag{4.23}$$

Inequalities (i) and (iii) follow from Proiposition 4.6.6 below. Inequality (ii) follows from the optimality of γ^{POP} and inequality (iv) follows from the optimality of γ^{LP}. Therefore, we obtain:

$$1 = \frac{POP(\gamma^{LP})}{POP(\gamma^{LP})} \leq \frac{POP(\gamma^{POP})}{POP(\gamma^{LP})} \leq \frac{LP(\gamma^{LP}) + \overline{R}}{POP(\gamma^{LP})} \leq \frac{POP(\gamma^{LP}) + \overline{R}}{POP(\gamma^{LP})} = 1 + \frac{\overline{R}}{POP(\gamma^{LP})}.$$

□

The proof of Theorem 4.6.5 relies on the following result.

Proposition 4.6.6. *For a given promotion profile γ, with the promotion set:*

$$\{(t_1, k_1), \ldots, (t_n, k_n)\},$$

the POP profits can be written as follows:

$$POP(\gamma_{\{(t_1,k_1),\ldots,(t_n,k_n)\}}) = LP(\gamma_{\{(t_1,k_1),\ldots,(t_n,k_n)\}}) + ER(\gamma_{\{(t_1,k_1),\ldots,(t_n,k_n)\}}). \qquad (4.24)$$

Here, $ER(\gamma_{\{(t_1,k_1),\ldots,(t_n,k_n)\}})$ *represents the error term between the POP and the LP objectives and is given by:*

$$ER(\gamma_{\{(t_1,k_1),\ldots,(t_n,k_n)\}}) = \sum_{i=1}^{n} \sum_{j=i+1}^{n} (q^{k_j} - q^0) g_{t_j - t_i}(q^{k_i}). \qquad (4.25)$$

Consequently, for any feasible promotion profile γ*, the POP profits satisfies:*

$$LP(\gamma) \leq POP(\gamma) \leq LP(\gamma) + \overline{R}.$$

The proof of Proposition 4.6.6 can be found in Appendix 4.D. Proposition 4.6.6 states that the POP profits can be written as the sum of the LP approximation evaluated at the same promotion profile, plus some given error term that depends on the price differences and the functions $g_k(\cdot)$.

We next show that the POP profits are supermodular in promotions.

Corollary 1 (Supermodularity of POP profits in promotions). *Let*

$$A = \{(t_1, k_1), \ldots, (t_N, k_N)\}$$

be a set of promotions with $1 \leq t_1 < t_2 < \cdots < t_n$ *(*$n \leq L$*) and let* $B \subset A$*. Consider a new promotion* (t', k') *where* $t' \notin \{t_n\}_{n=1}^{N}$*. Then, the new promotion* (t', k') *yields a greater marginal increase in profits when added to* A *than when added to* B*, that is:*

$$POP(\gamma_{A\cup\{(t',k')\}}) - POP(\gamma_A) \geq POP(\gamma_{B\cup\{(t',k')\}}) - POP(\gamma_B). \qquad (4.26)$$

Proof. We first introduce the following definition. For two promotions (t, k) and (u, ℓ)

with $t \neq u$, we define the interaction function:

$$\phi((t,k),(u,\ell)) = \begin{cases} (q^\ell - q^0)g_{u-t}(q^\ell) & \text{if } u > t; \\ (q^k - q^0)g_{t-u}(q^k) & \text{if } t > u. \end{cases}$$

Since $q^k, q^\ell \leq q^0$, and $g_m(p) \leq 0$ for all m and p, we have $\phi((t,k),(u,\ell)) \geq 0$. Observe that:

$$POP(\gamma_{\{(t,k)\}}) = POP(\gamma^0) + b_t^k,$$

where b_t^k are defined in (4.6) and represent the unilateral deviations in total profits by applying a single promotion at time t with price q^k. Similarly, we have: $POP(\gamma_{\{(u,\ell)\}}) = POP(\gamma^0) + b_u^\ell$. Therefore, we obtain:

$$POP(\gamma_{\{(t,k),(u,\ell)\}}) = POP(\gamma_{\{(t,k)\}}) + POP(\gamma_{\{(u,\ell)\}}) - POP(\gamma^0) + \phi((t,k),(u,\ell)).$$

In other words, the function $\phi((t,k),(u,\ell))$ compensates for the interaction term when we do both promotions (t,k) and (u,ℓ) simultaneously. From equation (4.24) in Proposition 4.6.6, we obtain:

$$POP(\gamma_A) = LP(\gamma_A) + \sum_{(t,k),(u,\ell) \in A: t<u} (q^\ell - q^0)g_{u-t}(q^\ell)$$

$$POP(\gamma_{A \cup \{(t',k')\}}) = LP(\gamma_{A \cup \{(t',k')\}}) + \sum_{(t,k),(u,\ell) \in A \cup \{(t',k')\}: t<u} (q^\ell - q^0)g_{u-t}(q^\ell),$$

and similarly for the set B. By using the definition of the LP objective function:

$$LP(\gamma_{\{(t_1,k_1),\ldots,(t_n,k_n)\}}) = POP(\gamma^0) + \sum_{i=1}^{n} (POP(\gamma_{\{t_i,k_i\}}) - POP(\gamma^0)),$$

we obtain: $LP(\gamma_{A \cup \{(t',k')\}}) - LP(\gamma_A) = POP(\gamma') - POP(\gamma^0)$ and: $LP(\gamma_{B \cup \{(t',k')\}}) - LP(\gamma_B) = POP(\gamma') - POP(\gamma^0)$, where we define $\gamma' = \gamma_{\{(t',k')\}}$. One can now obtain the following relations:

$$POP(\gamma_{A\cup\{(t',k')\}}) - POP(\gamma_A) = POP(\gamma') - POP(\gamma^0) + \sum_{(t,k)\in A} \phi((t,k),(t',k')),$$

$$POP(\gamma_{B\cup\{(t',k')\}}) - POP(\gamma_B) = POP(\gamma') - POP(\gamma^0) + \sum_{(t,k)\in B} \phi((t,k),(t',k')).$$

Therefore, we obtain:

$$\begin{aligned}\Big(POP(\gamma_{A\cup\{(t',k')\}}) - POP(\gamma_A)\Big) - \Big(POP(\gamma_{B\cup\{(t',k')\}}) - POP(\gamma_B)\Big)& \\ = \sum_{(t,k)\in A\setminus B} \phi((t,k),(t',k')) \geq 0.&\end{aligned}$$

□

Corollary 1 states that for an additive demand model as in (4.18), the POP profits are supermodular in promotions. Note that unlike in the multiplicative case, the claim is valid for any set of promotions. Consequently, it supports intuitively the fact that the LP approximation underestimates the POP objective, i.e., $POP(\gamma^{POP}) \geq LP(\gamma^{POP})$. Note that by considering the objective (total profits) of problem (POP) as a continuous function of the prices $p_1, p_2, \ldots, p_T$, one can equivalently show the supermodularity property by checking the non-negativity of all the cross-derivatives. We next show that the upper and lower bounds of Theorem 4.6.5 are tight.

Proposition 4.6.7 (Tightness of the bounds for additive model).

1. *The lower bound in Theorem 4.6.5 is tight. More precisely, for any given price ladder, L, S and functions g_k, there exist T, costs c_t and functions f_t such that:*

$$POP(\gamma^{POP}) = POP(\gamma^{LP}).$$

2. *The upper bound in Theorem 4.6.5 is tight. More precisely, for any given price*

ladder, L, S and functions g_k, there exist T, costs c_t and functions f_t such that:

$$POP(\gamma^{POP}) = POP(\gamma^{LP}) + \overline{R}.$$

The proof can be found in Appendix 4.E.

Illustrating the bounds.

For brevity, the plots where we illustrate the bounds for the additive demand model are presented in Appendix 4.F. We refer the reader to Section 4.6.1 for a discussion of the plots as a function of the various parameters since the trends we observe are similar in both the multiplicative and additive models.

4.6.3 Unified Model

In this section, we consider a unified demand model that has both multiplicative and additive components. In other words, the past prices have simultaneously a multiplicative and an additive effect on current demand:

$$d_t = \lambda \cdot d_1(p_t, p_{t-1}, \ldots, p_{t-M}) + (1 - \lambda) \cdot d_2(p_t, p_{t-1}, \ldots, p_{t-M}), \tag{4.27}$$

where $d_1(p_t, p_{t-1}, \ldots, p_{t-M})$ is a multiplicative model as in (4.13) and

$$d_2(p_t, p_{t-1}, \ldots, p_{t-M})$$

is an additive model as in (4.18). The parameter $0 \leq \lambda \leq 1$ represents the fraction of the demand that behaves according to the multiplicative demand model. This model in (4.27) can be used to capture a pool of consumers with different segments identified from data. More specifically, the consumers can be partitioned into segments, such as loyal and non-loyal members. In this case, λ is calibrated depending on the proportion of the appropriate segment. It is likely that the demand estimation for the various segments yields different demand models and one can then combine them into an aggregate form as in (4.27). Note that if $\lambda = 0$, (4.27) reduces to the additive class of demand functions we discussed in Section 4.6.2; whereas if $\lambda = 1$, (4.27) reduces to the multiplicative class of demand functions we discussed in Section 4.6.1. We also note that this approach can be extended to include more than two segments depending on the context and on the data available.

In order to solve the POP for the case with the unified demand model in (4.27), one can still naturally use the LP approximation method described in Section 4.5. However, the guarantees relative to the optimal profits we have shown are valid only for the multiplicative or the additive demand forms (i.e., when either $\lambda = 0$ or 1). Our goal is to extend the bounds on the quality of the LP approximation for the unified demand model in (4.27). We note that for the unified demand model in (4.27), the resulting POP is generally neither submodular nor supermodular in the promotions. Consequently, it is not easy to solve such problems to optimality and even getting a good approximation solution can be challenging. We next show that our LP based solution still yields a good approximation along with the lower and upper bounds.

Consider the following three solutions: γ^{LP_1} , γ^{LP_2} and $\gamma^{LP_{unif}}$ that correspond to the LP approximation of the multiplicative, additive and unified demand models respectively. We denote:

$$\Pi = \max\left\{POP_1(\gamma^{LP_1}), POP_2(\gamma^{LP_2}), POP(\gamma^{LP_{unif}})\right\}, \tag{4.28}$$

where $POP_1(\gamma^{LP_1})$ ($POP_2(\gamma^{LP_2})$) corresponds to the POP objective function for the additive (multiplicative) part of the demand only, i.e., $\lambda = 0$ ($\lambda = 1$) evaluated at the corresponding LP approximation solution. Since the three solutions in (4.28) are feasible to the POP for the unified demand model, we obtain:

$$\Pi \leq POP(\gamma^{POP_{unif}}), \tag{4.29}$$

where $\gamma^{POP_{unif}}$ corresponds to the optimal POP solution for the unified demand model. The bounds of the LP approximation relative to the optimal POP solution for the unified demand model in (4.27) are presented in the following Theorem.

Theorem 4.6.8. *Let $\gamma^{POP_{unif}}$ be an optimal solution to* (POP), *and let* Π *be defined as in* (4.28)*. Then:*

$$1 \leq \frac{POP(\gamma^{POP_{unif}})}{\Pi} \leq UB2 = \frac{\lambda}{\underline{R_1}} + (1-\lambda) \cdot \left[1 + \frac{\overline{R_2}}{POP_2(\gamma^{LP_2})}\right], \tag{4.30}$$

where $\underline{R_1}$ and $\overline{R_2}$ are given by (4.35) *and* (4.22) *respectively.*

Proof. The first inequality follows directly from equation (4.29). We next show the second inequality. First, we observe that the POP objective function for the unified demand model can be written as follows:

$$POP(\gamma^{POP_{unif}}) = \lambda \cdot POP_1(\gamma^{POP_{unif}}) + (1-\lambda) \cdot POP_2(\gamma^{POP_{unif}}),$$

where $POP_1(\gamma^{POP_{unif}})$ and $POP_2(\gamma^{POP_{unif}})$ represent the POP objective when the demand is multiplicative and additive respectively evaluated at the optimal solution of the POP for the unified model. By the optimality of POP_1 and POP_2, we have that:

$$\lambda \cdot POP_1(\gamma^{POP_{unif}}) + (1-\lambda) \cdot POP_2(\gamma^{POP_{unif}}) \leq \lambda \cdot POP_1(\gamma^{POP_1}) + (1-\lambda) \cdot POP_2(\gamma^{POP_2}).$$

By using the respective bounds for the multiplicative and additive demand models, we obtain:

$$\begin{aligned} &\lambda \cdot POP_1(\gamma^{POP_1}) + (1-\lambda) \cdot POP_2(\gamma^{POP_2}) \leq \\ &\qquad \frac{\lambda}{\underline{R_1}} \cdot POP_1(\gamma^{LP_1}) + (1-\lambda) \cdot \left[1 + \frac{\overline{R_2}}{POP_2(\gamma^{LP_2})}\right] \cdot POP_2(\gamma^{LP_2}). \end{aligned}$$

The proof can be concluded by using the definition of Π. □

We note that the upper bound is based on solving the demand segments separately and reduces to the special cases of additive and multiplicative demand when λ equals 0 and 1 respectively. Finally, we present an alternative bound in terms of the objective of the LP approximation problem.

Corollary 2. *Let $\gamma^{POP_{unif}}$ be an optimal solution to* (POP), *and let $\gamma^{LP_{unif}}$ be an optimal solution to the LP approximation. Then:*

$$POP(\gamma^{LP_{unif}}) \leq POP(\gamma^{POP_{unif}}) \leq UB1 = \lambda \cdot LP_1(\gamma^{LP_1}) + (1-\lambda) \cdot \left[LP_2(\gamma^{LP_2}) + \overline{R_2}\right], \tag{4.31}$$

where $\overline{R_2}$ is given by (4.22).

Proof. The first inequality follows from the feasibility of the LP solution. We next show the second inequality. The POP objective for the unified demand model can be written as follows:

$$POP(\gamma^{POP_{unif}}) = \lambda \cdot POP_1(\gamma^{POP_{unif}}) + (1-\lambda) \cdot POP_2(\gamma^{POP_{unif}}),$$

where $POP_1(\gamma^{POP_{unif}})$ $(POP_2(\gamma^{POP_{unif}}))$ represent the POP objective for the multiplicative (additive) segment exclusively evaluated at the optimal solution of the POP for the unified model. The optimality of POP_1 and POP_2 implies that:

$$\lambda \cdot POP_1(\gamma^{POP_{unif}}) + (1-\lambda) \cdot POP_2(\gamma^{POP_{unif}}) \leq \\ \lambda \cdot POP_1(\gamma^{POP_1}) + (1-\lambda) \cdot POP_2(\gamma^{POP_2}).$$

By using the respective bounds for the multiplicative and additive demand models, we obtain:

$$\lambda \cdot POP_1(\gamma^{POP_1}) + (1-\lambda) \cdot POP_2(\gamma^{POP_2}) \leq \\ LP_1(\gamma^{LP_1}) + (1-\lambda) \cdot \left[LP_2(\gamma^{LP_2}) + \overline{R_2}\right].$$

$\square$

In conclusion, by using the LP solution $\gamma^{LP_{unif}}$, one can obtain a feasible solution for the POP efficiently. In addition, for the unified demand model in (4.27) one can compute guarantees on the performance given in equation (4.31) even though the problem is generally neither submodular nor supermodular. This upper bound is obtained by solving the LP approximation separately for each segment of the demand and provides a certificate on the quality of the approximation. We will illustrate both upper bounds $UB1$ and $UB2$ in Appendix 4.G. This approach can be useful when several segments of consumers are identified from the data and can be viewed as a unifying framework of the multiplicative and additive demand models in Sections 4.6.1 and 4.6.2 respectively.

4.7 Computational Results

In order to quantify the value of our promotion optimization model, we perform an end-to-end experiment where we start with data from an actual retailer (supermarket), estimate the demand model we introduce, validate it, compute the optimized prices from our LP model and finally compare them with actual prices implemented by the retailer. In this section, following the recommendation of our industry collaborators, we perform detailed computational experiments for the log-log demand, which is a special case of the multiplicative model (4.13) and often used in practice.

4.7.1 Estimation Method

We obtained customer transaction data from a grocery retailer. The structure of the raw data is the customer loyalty card ID (if applicable), a timestamp, and the purchased items during that trip. In this paper, we focus on the coffee category at a particular store. For the purposes of demand estimation, we first aggregated the sales at the brand-week level. It seems natural to aggregate sales data at the week level as we observe that typically, a promotion starts on a Monday and ends on the following Sunday. Our data consists of 117 weeks from 2009 to 2011. For ease of interpretation and to keep the prices confidential, we normalize the regular price of each product to 1.

To predict demand as a function of prices, we estimate a log-log (power function) demand model incorporating seasonality and trend effects (similarly as in (4.10)):

$$\log d_{it} = \beta^0 BRAND_i + \beta^1 t + \beta^2 WEEK_t + \sum_{m=0}^{M} \beta^3_{im} \log p_{i,t-m} + \epsilon_t, \tag{4.32}$$

where i and t denote the brand and time indices, d_{it} denotes the sales (which we assume is equal to the demand, as we discussed in Section 4.3.2) of brand i in week t, $BRAND_i$ and $WEEK_t$ denote brand and week indicators, p_{it} denotes the average per-unit selling price of brand i in week t. β_0 and β_2 are vectors with components for each brand and each week respectively, whereas β_1 is a scalar that captures the trend. Note that the seasonality parameters β_2 for each week of the year are jointly estimated across all the brands in the category. The additive noises ϵ_t; $\forall t = 1, \ldots, T$ account for the unobserved discrepancies and are assumed to be normally distributed and i.i.d. Similar demand models have been used in the literature, e.g., Heerde et al. [2000] and Macé and Neslin [2004].

The model in (4.32) is a multiplicative model, which assumes that the brands share a common multiplicative seasonality; but each brand depends only on its own current and past prices; and the independent variables are assumed to have multiplicative effects on demand. In particular, the model incorporates a trend effect β^1, weekly seasonality β^2, and price effects β^3. When the memory parameter $M = 0$, then only the current price affects the demand in week t. When the memory parameter $M = 2$, then the demand in week t depends not only on the price in the current week p_t but also in the price of the two previous weeks p_{t-1} and p_{t-2}. We note that our model does not account explicitly for cross-brand effects, i.e., we assume that the demand for brand i depends only on the prices of brand i. This assumption is reasonable for certain products such as coffee because people are loyal about the brand they consume and do not easily switch between brands. In addition, the high predictive accuracy of our model validates this assumption.

For ease of notation, from this point, we drop the brand index i since we estimate and optimize for a single item model. Observe that one can define:

$$f_t(p_t) = \exp\left(\beta^0 + \beta^1 t + \beta^2 WEEK_t + \beta_0^3 \log p_t\right),$$
$$g_m(p_{t-m}) = \left(p_{t-m}\right)^{\beta_m^3}, \qquad m = 1, \ldots, M,$$

and therefore, equation (4.32) is in fact a special case of the multiplicative demand model in (4.13).

Based on our intuition, one expects to find the following from the estimation:

1. Since demand decreases as the current price increases, we would expect that the self-elasticity parameter is negative, i.e., $\beta_0^3 < 0$.

2. Since a deeper past promotion leads to a greater reduction in current demand, we would expect that the past elasticity parameters are positive, i.e., $\beta_m^3 \geq 0$ for $m > 0$.

3. Holding the depth of promotion constant, a more recent promotion leads to a greater reduction in current demand than the same promotion earlier in time. Therefore, we would expect that the past-elasticity parameters are decreasing in time, i.e., $\beta_m^3 > \beta_{m+1}^3$, for $m = 1, \ldots, M-1$.

We note that the conditions above are a special case of Assumption 11 for the log-log demand.

We divide the data into a training set, which comprises the first 82 weeks and a test set which comprises the second 35 weeks. We use the training set to estimate the demand model and then predict the out-of-sample sales to test our predictions. In order to measure forecast accuracy, we use the following forecast metrics. In the sequel, we use the notation s_t for the actual sales (or equivalently demand) and $\hat{s}_t$ be the forecasted values.

- The *mean absolute percentage error* (MAPE) is given by:

$$\text{MAPE} = \frac{1}{T}\sum_{t=1}^{T}\frac{|s_t - \hat{s}_t|}{s_t}.$$

The MAPE captures the average relative forecast error in absolute value. If the forecast is perfect, then the MAPE is equal to zero.

- The R^2 is given by:

$$R^2 = 1 - \frac{SS_{res}}{SS_{tot}},$$

where $\bar{s} = \sum_{t=1}^{T} s_t/T$, $SS_{tot} = \sum_{t=1}^{T}(s_t - \bar{s})^2$ and $SS_{res} = \sum_{t=1}^{T}(s_t - \hat{s}_t)^2$. We distinguish between in-sample (IS) and out-of-sample (OOS) R^2. If the forecast is perfect, then $R^2 = 1$. In addition, one can consider the adjusted R^2 as it is common in demand estimation. The latter adjusts the regular R^2 to account for the number of explanatory variables in the model relative to the number of data points available and is given by:

$$R^2_{adj} = 1 - (1 - R^2)\cdot\frac{n-1}{n-p-1},$$

where p is the total number of independent variables in the model (not counting the constant term), and n is the sample size.

- The *revenue bias* is measured as the ratio of the forecasted to actual revenue, and is given by:

$$\text{revenue bias} = \frac{\sum_{t=1}^{T} p_t\hat{s}_t}{\sum_{t=1}^{T} p_t s_t}.$$

4.7.2 Estimation Results and Discussion

Coffee Category

The coffee category is an appropriate candidate to test our model as it is common in promotion applications (see e.g., Gupta [1988] and Villas-Boas [1995]). We use a linear regression to estimate the parameters of the demand model in equation (4.32) for five different brands of coffee. For conciseness, we only present a subset of the estimation results for two coffee brands in Table 4.2. We compare the actual and predicted sales in Figure 4-7. Remember that our data consists of 117 weeks which we split into 82 weeks on training and 35 weeks of testing.

	Variable	Coefficient	Std Error	p-value
	$\log p_t$	-3.277	0.231	2e-16***
Brand1	$\log p_{t-1}$	0.518	0.229	0.024*
	$\log p_{t-2}$	0.465	0.231	0.045*
	$\log p_t$	-4.434	0.427	2e-16***
Brand2	$\log p_{t-1}$	1.078	0.423	0.011*
	$\log p_{t-2}$	0.067	0.413	0.870
	MAPE		0.116	
Brand1	OOS R^2		0.900	
	Revenue bias		1.059	
	MAPE		0.097	
Brand2	OOS R^2		0.903	
	Revenue bias		1.017	

Table 4.2: Forecast Metrics for Two Brands of Coffee

Note. In-sample adjusted R^2 Significance codes: * indicates significance < 0.05, *** indicates significance < 0.001.

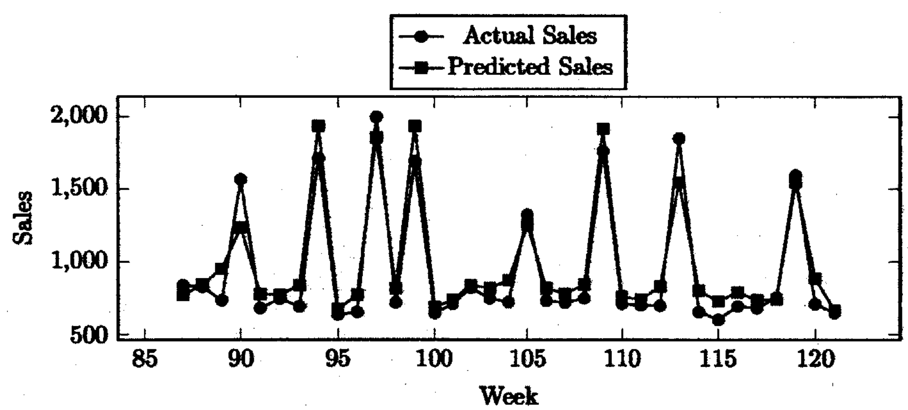


Figure 4-7: Demand Forecast for Brand 1 Test Set

Note. Plot of actual and forecasted sales over the 35 test weeks for Brand1.

On one hand, brand 1 is a private-label brand of coffee which has frequent promotions (approximately once every 4 weeks). The price-elasticity coefficients for the current price and two previous prices are statistically significant suggesting that for this brand, the memory parameter $M = 2$.

On the other hand, brand 2 is a premium brand of coffee which has also frequent promotions (approximately once every 5 weeks). The price-elasticity coefficients for both the current price and the price in the prior week are statistically significant, but the coefficient for the price two weeks ago is not. This suggests that for this brand, the memory parameter $M = 1$.

By observing the statistically significant price coefficients, one can observe that they agree with the expected findings mentioned previously. Furthermore, given the high accuracy as measured by low MAPEs, we expect that cross-brand effects are minimal.

Four Categories

In the same spirit, we estimate the log-log demand model for several brands for the chocolate, tea and yogurt categories. The results are summarized in Table 4.3. We do not report the individual product coefficients but note that they follow our expectations in terms of sign and ordering. We wish to highlight that the forecast error is low as evidenced by the high in-sample and out-of-sample R^2, and the low MAPE values and revenue bias being close to 1.

Category	IS Adj R^2	MAPE	OOS R^2	Revenue Bias	Product Memories
Coffee	0.974	0.115	0.963	1.000	0, 1, 2
Chocolate	0.951	0.185	0.872	0.990	0, 1, 2
Tea	0.984	0.187	0.759	1.006	0, 1
Yogurt	0.983	0.115	0.964	1.073	0, 1

Table 4.3: Forecast Metrics for Four Product Categories

We next observe the following regarding the effect of the memory parameter.

1. The memory parameter differs across products within a category. In general,

basic products have higher memory ($M = 1$ or 2) whereas premium items have lower memory ($M = 0$).

2. The memory parameters are estimated from data and differ depending on the category. Products in the yogurt and tea categories have memory of zero or one; whereas products in the coffee and chocolate categories have memory of zero, one or two. This agrees with our intuition that for perishable items (such as yogurt), consumers do not stock-pile and therefore, the memory parameter is zero. However, coffee is clearly a less perishable category so that stock-piling is more significant.

4.7.3 Optimization Results and Discussion

Having validated the forecasting demand model, we next perform a computational experiment to compute and test the optimized promotion prices. We assume that the demand forecast is the true demand model and use it as an input into our promotion optimization model from (POP).

Experimental setup: We compute the LP optimized prices for a single item (Brand1) over the horizon of the test weeks, which is $T = 35$ weeks. During the planning horizon, the retailer used $L = 8$ promotions with at most $S = 1$ separating weeks (i.e., consecutive promotions are separated by at least 1 week). As stated earlier, the regular price is normalized to be one unit. Due to confidentiality, we do not reveal the exact costs of the product, i.e., the parameters c_t in (POP). For the purpose of this experiment, we assume the cost of the product to be constant, $c_t = 0.4$. Since the lowest price charged by the retailer was 0.75, the set of permissible normalized prices is chosen to be $\{0.75, 0.80, 0.85, \ldots, 1\}$.

The LP optimization results are shown in Figure 4-8. Before discussing the results, we first want to make the following observations:

- The predicted profit using the actual prices implemented by the retailer (and not chosen optimally) together with the forecast model is \$18,425. All the results will be compared relative to this benchmark value.

- The predicted profit using only the regular price (i.e., no promotions) is \$17,890. This is a 2.9% loss relative to the benchmark. Therefore, the estimated log-log model predicts that the actual prices yield a 2.9% gain relative to the case without promotions, even if the actual promotions are not chosen optimally.

- The predicted profit using the optimized LP prices imposing the same number of promotions as a business requirement ($L = 8$ during a period of 35 weeks) is \$19,055. This is a 3.4% gain relative to the benchmark. Therefore, the estimated log-log model predicts that the optimized LP prices with the same number of promotions yield a 3.4% gain relative to the actual implemented profit. In other words, by only carefully planning the same number promotions, our model (and tool) suggests that the retailer can increase its profit by 3.4% in this case.

- The predicted profit using the optimized LP prices and allowing three additional promotions ($L = 11$) is \$19,362. This is a 5.1% gain relative to the benchmark. Therefore, the estimated log-log model predicts that optimized prices with three additional promotions yield a 5.1% gain relative to the actual profit. Therefore, the retailer can easily test the impact of allowing additional promotions within the horizon of $T = 35$ weeks.

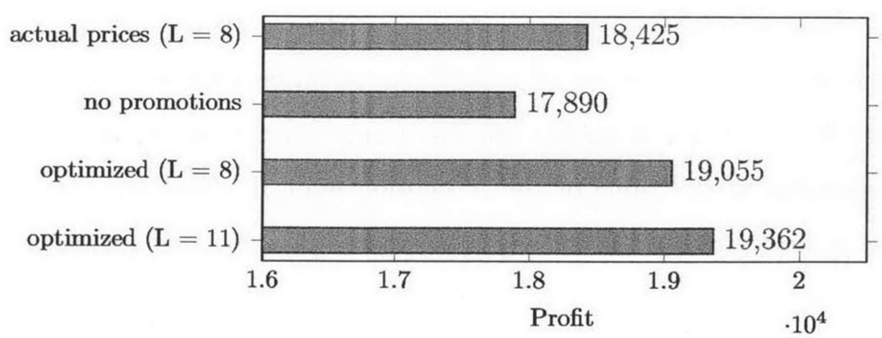


Figure 4-8: Predicted Revenue for Log-log Demand Model

We next compute the bound from Theorem 4.6.1 for the actual data we have been using in our computations above. The lower bound can be rewritten as $\underline{R} \cdot POP(\gamma^{POP}) \leq POP(\gamma^{LP})$, where $\underline{R} = \prod_{i=1}^{\tilde{L}-1} g_{i(S+1)}(q^K)$ and therefore, depends on the parameters of the problem. We compute $\underline{R}$ for both coffee brands from Table 4.2. We have $q^K = 0.75$, $L = 8$ and test the bound, $\underline{R}$, for various values of S. When $S \geq 2$, we observe that $\underline{R} = 1$ and therefore the method is optimal for both brands. For $S = 1$, we obtain that for Brand1, $\underline{R} = 0.8748$, whereas for Brand2, $\underline{R} = 1$. Finally, we consider $S = 0$ as it is the worst case scenario. In other words, no requirement on separating two successive promotions is imposed (not very realistic). We have for Brand1 and Brand2, $\underline{R} = 0.7538$ and $\underline{R} = 0.733$ respectively. We note that the above bounds outperform the approximation guarantees from the literature on submodular maximization. In particular, the problem of maximizing an arbitrary non-monotone submodular function subject to no constraints admits a $1/2$ approximation algorithm (see for example, Buchbinder et al. [2012] and Feige et al. [2011]). In addition, the problem of maximizing a monotone submodular function subject to a cardinality constraint admits a $1 - 1/e$ approximation algorithm (e.g., Nemhauser et al. [1978]). However, our bounds are not constant guarantees for every instance of the POP with multiplicative demand, as it depends on the values of the parameters. Recall also that in practice the LP approximation usually performs better than the bounds.

Next, we compare the running time of the LP to a naïve approach of using an exhaustive search method in order to find the optimal prices of the POP. Note that the POP problem is neither convex nor concave. The results are shown in Figure 4-9. The experiments were run using a desktop computer with an Intel Core i5 680 @ 3.60GHz CPU with 4 GB RAM. The LP formulation requires 0.01–0.05 seconds to solve, regardless of the value of the promotion limit L. However, the exhaustive search running time grows exponentially in L. In addition, for a simple instance of the problem with only 2 prices in the price ladder, it requires one minute to solve when $L = 8$. The running time of the exhaustive search method also grows exponentially in the number of elements of the price ladder. For example, with 3 elements in the price ladder and $L = 8$, it requires 3 hours to solve, whereas the LP solution solves within milliseconds. We note that since we are considering non-linear demand functions with integer variables, general methods to solve this problem do not exist in commercial solvers and hence are not practical.

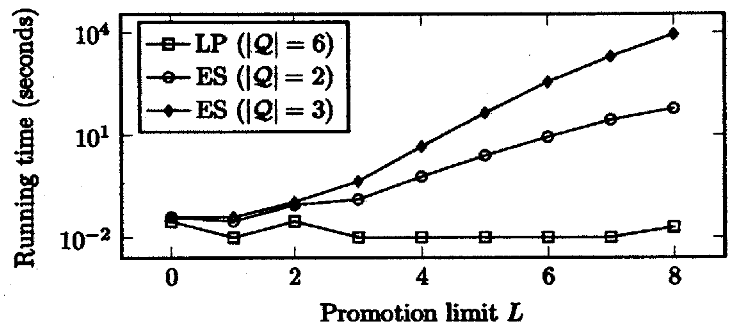


Figure 4-9: LP and Exhaustive Search Running Times

Note. **Running times of the LP formulation and the exhaustive search method for the POP. The number of the price ladder elements is $|Q|$. The LP was solved using the Java interface to Gurobi 5.5.0.**

4.8 Conclusions

In many important settings, promotions are a key instrument for driving sales and profits. We introduce and study an optimization formulation for the POP that captures several important business requirements as constraints (such as separating periods and promotion limits). We propose two general classes of demand functions depending on whether past prices have a multiplicative or an additive effect on current demand. These functions capture the promotion fatigue effect emerging from the stock-piling behavior of consumers and can be easily estimated from data. We show that for multiplicative demand, promotions have a supermodular effect (for some subsets of promotions) which leads to the LP approximation being an upper bound on the POP objective; whereas for additive demand, promotions have a submodular effect which leads to the LP approximation being a lower bound on the POP objective. The objective is nonlinear (neither convex nor concave) and the feasible region has linear constraints with integer variables. Since the exact formulation is "hard", we propose a linear approximation that allows us to solve the problem efficiently as an LP by showing the integrality of the IP formulation. We develop analytical results on the LP approximation accuracy relative to the optimal (but intractable) POP solution and characterize the bounds as a function of the problem parameters. We also show computationally that the formulation solves fast using actual data from a grocery retailer and that the accuracy is high.

Together with our industry collaborators from Oracle Retail, our framework allows us to develop a tool which can help supermarket managers to better understand promotions. We test our model and solution using actual sales data obtained from a supermarket retailer. For four different product categories, we estimate from transactions data the log-log and linear demand models (the linear model is relegated to the Appendix). Our estimation results provide a good fit and explain well the data but also reveal interesting insights. For example, non-perishable products exhibit longer memory in the sense that the sales are affected not only by the current price but also by the past prices. This observation validates the hypothesis that demand exhibits a promotion fatigue effect for certain items. We test our approach for solving the promotion optimization problem, by first estimating the demand model from data. We then solve the POP by using our LP approximation method. In this case, using the LP optimized prices would lead about 3% profit gain for the retailer, with even 5% profit gain by slightly modifying the number of promotions allowed. In addition, the running time of our LP is short (~0.05 seconds) making the method attractive and efficient. The naïve optimal exhaustive search method is several orders of magnitude slower. The fast running time allows the LP formulation to be used interactively by a category manager who may manage around 300 SKUs in a category. In addition, one can conveniently run a large number of instances allowing to perform a comprehensive sensitivity analysis translated into "what-if" scenarios. We are currently in the process of conducting a pilot experiment with an actual retailer, where we test our model in a real-world setting by optimizing promotions for several items and stores.

Acknowledgments

We would like to thank Oracle Corporation for their generous support through an ERO grant that made this work possible. This work was also partly supported by NSF grant CMMI-1162034. We would also like to thank Steve Jeffreys from Oracle Retail for his suggestions and support. In addition, we would like to thank Andrew Vakhutinsky and Peter Gaidarev formerly from Oracle Retail for their valuable feedback in this work. We would also like to thank Steve Graves, Adam Mersereau, Paat Rusmevichientong, Chung-Piaw Teo and Huseyin Topaloglu for their insightful comments that have helped improve this paper. Finally, we would like to thank the participants of the 2013 London Business School Collaborative Academic/Practitioner Workshop on Operational Innovation for their valuable feedback and discussions.

Appendix

4.A Proof of Lemma 4.6.2

Proof. 1. Since the proof may not be easy to follow, we present it together with a concrete example to illustrate the different steps. Let $T = 6$, $q^0 = 7$, $A = \{(1,1),(3,3)\}$, $B = \{(3,3)\}$ and $(t',k') = (5,5)$. We denote by $POP_t(\mathbf{p}_A)$ the profits at time t for the price vector $\mathbf{p}_A$. In addition, we further assume that: $\delta_5 = g_4(1) = 0.8$, $\delta_6 = g_5(1) = 0.9$. We next define the following quantities:

$$\begin{aligned}
a_t &= POP_t(\mathbf{p}_A) &&= POP_t(1,7,3,7,7,7)\\
a'_t &= POP_t(\mathbf{p}_{A\cup(t',k')}) &&= POP_t(1,7,3,7,5,7)\\
b_t &= POP_t(\mathbf{p}_B) &&= POP_t(7,7,3,7,7,7)\\
b'_t &= POP_t(\mathbf{p}_{B\cup(t',k')}) &&= POP_t(7,7,3,7,5,7).
\end{aligned}$$

For each time t, we define the following coefficient:

$$\delta_t = \frac{g_1((\mathbf{p}_A)_{t-1}) \cdot g_2((\mathbf{p}_A)_{t-2}) \cdots g_{t-1}((\mathbf{p}_A)_1)}{g_1((\mathbf{p}_B)_{t-1}) \cdot g_2((\mathbf{p}_B)_{t-2}) \cdots g_{t-1}((\mathbf{p}_B)_1)}.$$

δ_t represents the multiplicative reduction in demand at time t from the promotions present in the set A but not in B. Observe that from Assumption 11, we have $o \leq \delta_{t'} \leq \delta_{t'+1} \leq \cdots \leq \delta_T \leq 1$. In addition, we have: $a_t = \delta_t b_t$, $a'_t = \delta_t b'_t$. Observe also that condition (4.16) is equivalent to:

$$\sum_{t=1}^{T} a'_t - \sum_{t=1}^{T} a_t \geq 0. \tag{4.33}$$

Note that $a_t = a'_t$ for all $t < t'$. In the example, we have $a_1 = a'_1, \ldots, a_4 = a'_4$ as the prices in periods 1-4 are the same. Therefore, (4.33) becomes: $\sum_{t=t'}^{T} a'_t \geq \sum_{t=t'}^{T} a_t$. In the example, we obtain: $a'_5 + a'_6 \geq a_5 + a_6$. Note that $a'_t \leq a_t$ for any $t > t'$. In the example, a'_t has a promotion at $t = 5$ However, there is no promotion in a_t at $t = 5$ and therefore, the objective at $t = 6$ for a'_t is lower than the one in a_t, i.e., $a'_6 \leq a_6$. This implies that:

$$a'_{t'} - a_{t'} \geq \sum_{t=t'+1}^{T} (a_t - a'_t) \geq 0.$$

In the example, this translates to $a'_5 - a_5 \geq a_6 - a'_6 \geq 0$. We next multiply the left hand side by $1/\delta_{t'}$ and the terms in the right hand side by $1/\delta_t$ (recall that $1/\delta_{t'} \geq 1/\delta_t$ for $t > t'$). Therefore, we obtain:

$$b'_{t'} - b_{t'} = \frac{a'_{t'} - a_{t'}}{\delta_{t'}} \geq \sum_{t=t'+1}^{T} \left(\frac{a_t - a'_t}{\delta_t}\right) = \sum_{t=t'+1}^{T} (b_t - b'_t) \geq 0.$$

In the example, this translates to: $b'_5 - b_5 = \frac{a'_5 - a_5}{0.8} \geq \frac{a_6 - a'_6}{0.9} = b_6 - b'_6 \geq 0$. Recall that our goal is to show equation (4.17), or alternatively: $\sum_{t=1}^{T} a'_t - \sum_{t=1}^{T} a_t \leq \sum_{t=1}^{T} b'_t - \sum_{t=1}^{T} b_t$. Note that this is equivalent to: $\sum_{t=t'}^{T}(a'_t - a_t) = \sum_{t=t'}^{T} \delta_t(b'_t - b_t) \leq \sum_{t=t'}^{T}(b'_t - b_t)$. By rearranging the terms, we obtain:

$$\sum_{t=t'+1}^{T} (1 - \delta_t)(b_t - b'_t) \leq (1 - \delta_{t'})(b'_{t'} - b_{t'}).$$

In the example, this would be: $0.1(b_6 - b_6') \leq 0.2(b_5' - b_5)$. Finally, note that the above inequality is true because of the following:

$$\sum_{t=t'+1}^{T} (1-\delta_t)(b_t - b_t') \leq \sum_{t=t'+1}^{T} (1-\delta_{t'})(b_t - b_t') \leq (1-\delta_{t'})(b_{t'}' - b_{t'}).$$

In the example, this is clear because: $0.1(b_6 - b_6') \leq 0.2(b_6 - b_6') \leq 0.2(b_5' - b_5)$.

2. We first introduce the following notation. Let γ^{POP} be an optimal solution to the POP and $\{(t_1, k_1), \ldots, (t_n, k_n)\}$ the set of promotions in γ^{POP}. For any subset $B \subset \{1, 2, \ldots, n\}$, we define: $\gamma(B) = \gamma(\{(t_i, k_i) : i \in B\})$. For example, let the price ladder be $\{q^0 = 5, q^1 = 4\}$ and $\gamma^{POP} = \gamma(\{(1,1),(3,1),(5,1)\})$. Then, $\gamma(\{1,3\}) = \gamma(\{(1,1),(5,1)\})$.

   Note that one can write the following telescoping sum:

$$POP(\gamma^{POP}) = POP(\gamma\{1\}) + \sum_{m=1}^{n-1} \Big[POP(\gamma\{1,\ldots,m+1\}) - POP(\gamma\{1,\ldots,m\})\Big].$$

   Based on Proposition 4.A.1 below, we have for each $m = 1, 2, \ldots n-1$:

$$POP(\gamma\{1,\ldots,m+1\}) - POP(\gamma\{1,\ldots,m\}) \geq 0.$$

   By applying the submodularity property from Lemma 4.6.2 part 1, we obtain: $0 \leq POP(\gamma\{1,\ldots,m+1\}) - POP(\gamma\{1,\ldots,m\}) \leq POP(\gamma\{m+1\}) - POP(\gamma^0)$. Therefore, we have:

$$\begin{aligned} POP(\gamma^{POP}) &= POP(\gamma\{1\}) + \sum_{m=1}^{n-1} \Big[POP(\gamma\{1,\ldots,m+1\}) - POP(\gamma\{1,\ldots,m\})\Big] \\ &\leq POP(\gamma^0) + \sum_{m=1}^{n} \Big[POP(\gamma\{m\}) - POP(\gamma^0)\Big] = LP(\gamma^{POP}). \end{aligned}$$

$\square$

Proposition 4.A.1. *Let $n \geq 2$ be an integer and γ^{POP} an optimal solution to the*

POP with n promotions. Then, $POP(\gamma\{1,\ldots,m+1\}) - POP(\gamma\{1,\ldots,m\}) \geq 0$ for $m = 1,2,\ldots,n-1$.

Proof. The proof proceeds by induction on the number of promotions. We first show that the claim is true for the base case i.e., $n = 2$. By the optimality of $\gamma^{POP} = \gamma\{1,2\}$, we have:

$$0 \leq POP(\gamma\{1,2\}) - POP(\gamma\{1,2\}).$$

Next, we assume that the claim is true for n and show its correctness for $n+1$. Let POP' denote the POP problem with the additional constraint that promotion (t_1, k_1) is used, i.e., $p_{t_1} = q^{k_1}$. One can see that the set of promotions $\{(t_2, k_2), \ldots, (t_{n+1}, k_{n+1})\}$ is an optimal solution to POP' with n promotions. Therefore, by using the induction hypothesis, we have:

$$\begin{array}{lll} POP'(\gamma\{2,\ldots,n,n+1\}) & -POP'(\gamma\{2,\ldots,n\}) & \geq 0 \\ \vdots & \vdots & \vdots \\ POP'(\gamma\{2,3\}) & -POP'(\gamma\{2\}) & \geq 0 \end{array}$$

Equivalently, in terms of the POP:

$$\begin{array}{lll} POP(\gamma\{1,\ldots,n,n+1\}) & -POP(\gamma\{1,\ldots,n\}) & \geq 0 \\ \vdots & \vdots & \vdots \\ POP(\gamma\{1,2,3\}) & -POP(\gamma\{1,2\}) & \geq 0 \end{array}$$

Therefore, it remains to show that: $POP(\gamma\{1,2\}) - POP(\gamma\{1\}) \geq 0$. We next prove the following chain of inequalities:

$$
\begin{aligned}
& POP(\gamma\{1,2\}) - POP(\gamma\{1\}) \\
\geq\ & POP(\gamma\{1,2,3\}) - POP(\gamma\{1,3\}) \\
\geq\ & POP(\gamma\{1,2,3,4\}) - POP(\gamma\{1,3,4\}) \\
\vdots\ & \\
\geq\ & POP(\gamma\{1,\ldots,n,n+1\}) - POP(\gamma\{1,3,4,\ldots,n,n+1\}).
\end{aligned} \tag{4.34}
$$

By using the induction hypothesis together with the submodularity property from Lemma 4.6.2 part 1, we obtain for each $m = 2,3,\ldots,n-1$:

$$
\begin{aligned}
POP(\gamma\{1,\ldots,m,m+1\}) - POP(\gamma\{1,\ldots,m\}) \leq \\
POP(\gamma\{1,3,4,\ldots,m,m+1\}) - POP(\gamma\{1,3,4,\ldots,m\}).
\end{aligned}
$$

Finally, from the optimality of $\gamma\{1,\ldots,n,n+1\}$ for the POP, we have:

$$POP(\gamma\{1,\ldots,n,n+1\}) - POP(\gamma\{1,\ldots,n\}) \geq 0.$$

By rearranging the terms in the above equations, one can derive the chain of inequalities in (4.34) and this concludes the proof. □

4.B Proof of Proposition 4.6.3

Proof. We denote the set of promotions in the price vector $\mathbf{p}$ by:

$$\mathbf{p} = \mathbf{p}\{(t_1, q^{t_1}), \ldots, (t_N, q^{t_N})\},$$

where N is the number of promotions. The price vector $\mathbf{p}^n = \mathbf{p}\{(t_n, q^{t_n})\}$ for each $n = 1, \ldots, N$ denotes the single promotion price at time t_n (no promotion at the remaining periods). By convention, let us denote $n = 0$ to be the regular price only vector $\mathbf{p}^0 = (q^0, \ldots, q^0)$. We denote the cumulative POP objective in periods $[u, v)$ when using $\mathbf{p}^n$ by:

$$x^n_{[u,v)} = POP(\mathbf{p}\{(t_n, q^{t_n})\})_{[u,v)} = \sum_{t=u}^{v-1} \mathbf{p}_t\{(t_n, q^{t_n})\} d_t(\mathbf{p}_t\{(t_n, q^{t_n})\}).$$

Note that the LP objective can be written as: $LP(\mathbf{p}) = x^0_{[1,T]} + \sum_{n=1}^{N} \left(x^n_{[1,T]} - x^0_{[1,T]} \right)$.

Since $\mathbf{p}^n$ and $\mathbf{p}^0$ do not promote before t_n, we have $x^n_{[1,t_n)} = x^0_{[1,t_n)}$. In addition, since $\mathbf{p}^n$ promotes at $t = t_n$ and $\mathbf{p}^0$ does not, the vector $\mathbf{p}^n$ yields a lower objective for the periods after t_n, i.e., $x^n_{[t_{n+1},T]} \leq x^0_{[t_{n+1},T]}$. Therefore, we obtain for each $n = 1, \ldots, N$:

$$\begin{aligned} x^n_{[1,T]} - x^0_{[1,T]} &= x^n_{[1,t_n)} + x^n_{[t_n,t_{n+1})} + x^n_{[t_{n+1},T]} - x^0_{[1,t_n)} - x^0_{[t_n,t_{n+1})} - x^0_{[t_{n+1},T]} \\ &\leq x^n_{[t_n,t_{n+1})} - x^0_{[t_n,t_{n+1})}. \end{aligned}$$

Therefore: $LP(\mathbf{p}) \leq UB = x^0_{[1,T]} + \sum_{n=1}^{N} \left(x^n_{[t_n,t_{n+1})} - x^0_{[t_n,t_{n+1})} \right) = x^0_{[1,t_1)} + \sum_{n=1}^{N} x^n_{[t_n,t_{n+1})}$. Let UB_t denote the value of UB at time t. Specifically, if $t \in [t_n, t_{n+1})$, then $UB_t = x^n_t$. We can write for any feasible price vector $\mathbf{p}$: $POP(\mathbf{p}) = \sum_{t=1}^{T} a_t UB_t$, where a_t is the decrease in demand at time t due to the past promotions in $\mathbf{p}$. In particular, if $t_n < t \leq t_{n+1}$, then: $a_t = g_{t-t_1}(q^{t_1}) g_{t-t_2}(q^{t_2}) \cdots g_{t-t_n}(q^{t_n})$. Since $0 \leq \underline{R} \leq a_t \leq 1$, we obtain: $\underline{R} \cdot LP(\mathbf{p}) \leq \underline{R} \cdot UB \leq POP(\mathbf{p})$. □

4.C Proofs of Tightness for Multiplicative Demand

1. **Lower bound**

Proof. In the case when $S \geq M$, we know from Proposition 4.5.1 that the LP approximation is exact. Therefore, the result holds in this case.

We next consider that $S < M$ and construct an instance of the POP as well as a price vector $\mathbf{p}^*$. We then show that this price vector $\mathbf{p}^*$ is optimal for both the POP and the LP approximation.

Let $T = L(M+1)$ and let us define the following price vector:

$$\mathbf{p}^* = \Big(q^K, \underbrace{q^0, \ldots, q^0}_{M \text{ times}}, q^K, \underbrace{q^0, \ldots, q^0}_{M \text{ times}}, \ldots, q^K, \underbrace{q^0, \ldots, q^0}_{M \text{ times}}\Big).$$

Let $\mathbf{U} = \{1, (M+1)+1, 2(M+1)+1, \ldots, (L-1)(M+1)+1\}$ denote the set of promotion periods in $\mathbf{p}^*$. We choose the demand functions f_t to be:

$$f_t(p_t) = \begin{cases} Z/q^K & \text{if } t \in \mathbf{U} \text{ and } p_t = q^K, \\ 1/q^0 & \text{otherwise,} \end{cases}$$

where:

$$Y = 1 + \sum_{m=1}^{M} (1 - g_m(q^K)),$$
$$Z = (M+2)Y.$$

We define all the costs to be zero, i.e., $c_t = 0, \forall t = 1, \ldots, T$. We prove the proposition by the following steps:

Step 1: We show that $\mathbf{p}^*$ is an optimal LP solution.

Step 2: We show that there exists an optimal POP solution with promotions only during periods $t \in \mathbf{U}$.

Step 3: We show that if $\mathbf{p}$ promotes only during periods $t \in \mathbf{U}$, then $POP(\mathbf{p}) \leq POP(\mathbf{p}^*)$.

By combining steps 2 and 3, we conclude that $\mathbf{p}^*$ is an optimal POP solution. Consequently, $POP(\mathbf{p}^{POP}) = POP(\mathbf{p}^{LP})$, implying that the lower bound is tight.

Proof of Step 1. By definition, we have: $POP(\mathbf{p}\{(t,K)\}) = POP(\mathbf{p}^0)+Z-Y$ for $t \in \mathbf{U}$. Therefore the LP coefficients as defined in (4.6) are given by:

$$b_t^k = \begin{cases} Z-Y & \text{if } t \in \mathbf{U}, k = K, \\ \leq 0 & \text{otherwise.} \end{cases}$$

Any LP optimal solution selects at most L of γ_t^k, for $k = 1, \ldots, K$ to be 1. Consequently, the optimal LP objective is bounded above by $T + L(Z-Y)$. In fact, the following γ^{LP} achieves this bound and is therefore optimal:

$$(\gamma^{LP})_t^k = \begin{cases} 1 & \text{if } t \in \mathbf{U}, k = K \\ 1 & \text{if } t \notin \mathbf{U}, k = 0 \\ 0 & \text{otherwise} \end{cases}$$

We then conclude that $\mathbf{p}^{LP} = \mathbf{p}^*$ is an optimal LP solution.

Proof of Step 2. Consider any feasible price vector $\mathbf{p}$ and let $\mathbf{A}$ be the set of promotions in $\mathbf{p}$. We next show that $POP(\mathbf{p}) \leq POP(\mathbf{p}^*)$ so that $\mathbf{p}^*$ is an optimal POP solution. If $\mathbf{p}$ uses the promotion $p_t = q^k$ during a period $t \notin \mathbf{U}$, then we can consider the reduced set of promotions $\mathbf{B} = \mathbf{A} \setminus \{(t,k)\}$. Note that the promotion (t,k) does not increase the profit at time t. Indeed, decreasing the price p_t will not increase the profit at time t since $f_t(p_t) = 1/q^0$ for all p_t, and potentially will reduce the profit in future periods $t+1, \ldots, t+M$. Thus, removing the promotion (t,k) increases the total profit, that is $POP(\gamma(\mathbf{A})) \leq POP(\gamma(\mathbf{B}))$ .By applying this procedure repeatedly, one can reach a price vector with only promotions in periods $t \in \mathbf{U}$ that achieves a profit at least equal to $POP(\mathbf{p})$. In other words, there exists an optimal POP solution with promotions only during periods $t \in \mathbf{U}$.

Proof of Step 3. Let $\mathbf{p}$ be a price vector that only contains promotions during periods $t \in \mathbf{U}$. Let n be the number of periods t in $\mathbf{p}$ such that $p_t = q^K$ ($n \leq L$ because $\mathbf{U}$ is composed of L periods). Note that all the successive promotions in $\mathbf{U}$ are separated by at least M periods so that each pair of promotions of $\mathbf{p}$ does not interact. Therefore, the profit of $\mathbf{p}$ is given by:

$$POP(\mathbf{p}) = POP(\mathbf{p}^0) + n(Z - Y) \leq POP(\mathbf{p}^0) + L(Z - Y).$$

From the definition of $\mathbf{p}^*$, we have that $POP(\mathbf{p}^*) = POP(\mathbf{p}^0) + L(Z-Y)$. Indeed, each promotion (t, K) of $\mathbf{p}^*$ results in an increase in profit of $Z - Y$, and each pair of promotions of $\mathbf{p}^*$ is separated by at least M periods so that there is no interaction between promotions. Consequently, $\mathbf{p}^*$ is an optimal POP solution and the lower bound is tight. □

2. **Upper bound**

Proof. Let us denote the bound with n promotions by:

$$\underline{R}_n = \prod_{i=1}^{n-1} g_{i(S+1)}(q^K), \tag{4.35}$$

when $\underline{R}_0 = 1$ by convention. We can also define the following limit:

$$\underline{R}_\infty = \lim_{n \to \infty} \underline{R}_n.$$

Note that $g_m(q^K) \leq 1$ so that R_n is non-increasing with respect to n. Note also that $g_m(q^K) = 1$ for $m > M$ so that $\underline{R}_{M+1} = \underline{R}_{M+2} = \cdots = \underline{R}_\infty$, i.e., the sequence $\underline{R}_n$ converges.

In the case when $S \geq M$, we know from Proposition 4.5.1 that the LP approximation is exact. We also know from (4.35) that $\underline{R}_n = 1$ for all n. Therefore, the result holds in this case.

We next consider that $S < M$ and define the following sequence of problems:

$$POP^n = POP(\langle q^k\rangle_{k=0}^K, \langle f_t^n\rangle_{t=1}^{T_n}, \langle c_t\rangle_{t=1}^{T_n}, \langle g_m\rangle_{m=1}^M, L_n, S),$$

where $\langle q^k\rangle_{k=0}^K, \langle g_m\rangle_{m=1}^M, S$ are given parameters and the costs $c_t = 0$. In addition, $L_n = n$, and $T_n = n(M+1)$. We choose the functions f_t^n to be equal:

$$f_t^n(p_t) = \begin{cases} Z/q^K & \text{if } 1 \le t \le LM+1 \text{ and } p_t = q^K, \\ 1/q^0 & \text{otherwise.} \end{cases}$$

where,

$$Y = 1 + \sum_{m=1}^{M}(1 - g_m(q^K)),$$

$$Z = 100Yn.$$

We prove the proposition by the following steps:

Step 1: We show that the following price vector is an optimal LP solution:

$$\mathbf{p}^{LP} = \Big(q^K, \underbrace{q^0, \ldots, q^0}_{S \text{ times}}, q^K, \underbrace{q^0, \ldots, q^0}_{S \text{ times}}, \ldots, q^K, \underbrace{q^0, \ldots, q^0}_{T-(L-1)(S+1)-1 \text{ times}}\Big).$$

Step 2: We show that:

$$POP^n(\mathbf{p}_n^{LP}) \le T - L + Z(\underline{R}_1 + \cdots + \underline{R}_n).$$

Step 3: We show the following lower bound for the optimal profit: $POP^n(\mathbf{p}_n^{POP}) \ge nZ$.

Step 4: We finally prove the convergence of the following limit, implying the desired result:

$$\lim_{n\to\infty} \frac{POP_n(\mathbf{p}_n^{POP})}{POP_n(\mathbf{p}_n^{LP})} = \frac{1}{\underline{R}_\infty}.$$

Proof of Step 1. Based on the above definitions, we have: $POP(\mathbf{p}\{(t,K)\}) = POP(\mathbf{p}^0) + Z - Y$ for $1 \le t \le LM + 1$. Therefore, the LP coefficients are given by:

$$b_t^k = \begin{cases} Z - Y & \text{if } 1 \le t \le LM+1, k = K, \\ \le 0 & \text{otherwise.} \end{cases}$$

Let $\mathbf{U} = \{1, S+1, 2S+1, \ldots, LS+1\}$ denote the set of promotion periods in $\mathbf{p}^{LP}$.

Any LP optimal solution selects at most L of γ_t^k, for $k = 1, \ldots, K$ to be 1. Consequently, the optimal LP objective is bounded above by $T + L(Z - Y)$. In fact, the following γ^{LP} achieves this bound and is therefore optimal:

$$(\gamma^{LP})_t^k = \begin{cases} 1 & \text{if } t \in \mathbf{U}, k = K \\ 1 & \text{if } t \notin \mathbf{U}, k = 0 \\ 0 & \text{otherwise} \end{cases}$$

Therefore, we conclude that the price vector $\mathbf{p}^{LP}$ is an optimal LP solution.

Proof of Step 2. One can see that the profit induced by the i-th promotion of $\mathbf{p}_L^{LP}$ (at time $t = (i-1)S + 1$) is $\underline{R}_i Z$ due to the effect of the promotions $1, 2, \ldots, (i-1)$. In addition, the profit from each non-promotion period is bounded above by 1. We obtain:

$$POP_n(\gamma_n^{LP}) \le T - L + Z(\underline{R}_1 + \underline{R}_2 + \cdots + \underline{R}_n).$$

Proof of Step 3. Consider the following price vector:

$$\mathbf{p} = \Big(q^K, \underbrace{q^0, \ldots, q^0}_{M \text{ times}}, q^K, \underbrace{q^0, \ldots, q^0}_{M \text{ times}}, \ldots, q^K, \underbrace{q^0, \ldots, q^0}_{M \text{ times}}\Big).$$

Note that $\mathbf{p}$ is feasible for POP_n. Note that all the successive promotions are separated by at least M periods so that each pair of promotions of $\mathbf{p}$ does not interact. Therefore, the profit induced by the i-th promotion in $\mathbf{p}$ (at time $t = (i-1)M+1$) is Z. As a result, we obtain the following lower bound for the POP profit of $\mathbf{p}$:

$$POP_n(\mathbf{p}) \geq nZ.$$

This also provides us a lower bound for the optimal POP profit:

$$POP_n(\mathbf{p}_n^{POP}) \geq POP_n(\mathbf{p}) \geq nZ.$$

Proof of Step 4. We show that $\frac{1}{\underline{R}_\infty}$ is both a lower and upper bound of the limit. First, using Theorem 4.6.1 for POP^n, we have:

$$\frac{POP^n(\mathbf{p}_n^{POP})}{POP^n(\mathbf{p}_n^{LP})} \leq \frac{1}{\underline{R}_n}.$$

By taking the limit when $n \to \infty$ on both sides:

$$\lim_{n\to\infty} \frac{POP^n(\mathbf{p}_n^{POP})}{POP^n(\mathbf{p}_n^{LP})} \leq \lim_{n\to\infty} \frac{1}{\underline{R}_n} = \frac{1}{\underline{R}_\infty}.$$

By using Steps 2 and 3, we obtain:

$$\begin{aligned}
\lim_{n\to\infty} \frac{POP^n(\mathbf{p}_n^{POP})}{POP^n(\mathbf{p}_n^{LP})} &\geq \lim_{n\to\infty} \frac{n \cdot 100nY}{nM + 100nY(\underline{R}_1 + \underline{R}_2 + \cdots + \underline{R}_n)} \\
&= \lim_{n\to\infty} \frac{1}{\frac{M}{100nY} + \frac{\underline{R}_1 + \underline{R}_2 + \cdots + \underline{R}_n}{n}} = \frac{1}{\underline{R}_\infty}.
\end{aligned}$$

In the last equality, we have used the fact that if $\langle a_n \rangle_{n=1}^{\infty}$ converges to a finite limit a, then $\langle \sum_{i=1}^{n} a_i/n \rangle_{n=1}^{\infty}$ also converges to a.

□

4.D Proof of Proposition 4.6.6

Proof. Without loss of generality, we consider the case with the costs equal to zero, i.e., $c_t = 0$; $\forall t$. We next show that both sides of equation (4.24) at each time period t are equal. Let us define the quantities $e_t^i = g_{t_i - t}(q^{k_i})$ for $t > t_i$ that capture the demand reduction at time t due to the earlier promotion q^{k_i} at time t_i. Let LP_t and POP_t denote the LP approximation and POP objectives at time t respectively. Consider a price vector of the form: $\mathbf{p}_{\{(t_1,k_1),\ldots,(t_n,k_n)\}}$. The LP approximation evaluated at this price vector is given by:

$$LP(\mathbf{p}_{\{(t_1,k_1),\ldots,(t_n,k_n)\}}) = POP(\mathbf{p}^0) + \sum_{i=1}^{n} \left[q^{k_i} POP(\mathbf{p}_{(t_i,k_i)}) - POP(\mathbf{p}^0) \right].$$

The POP objective using the single promotion (t_i, k_i) is given by:

$$POP(\mathbf{p}_{(t_i,k_i)}) = q^0 f_1(q^0) + \cdots + q^0 f_{t_i-1}(q^0) + q^{t_i} f_{t_i}(q^{k_i}) + \\ q^0 \left[f_{t_i+1}(q^0) + e_{t_i+1}^i \right] + \cdots + q^0 \left[f_T(q^0) + e_T^i \right].$$

In addition, we have: $POP(\mathbf{p}^0) = \sum_{t=1}^{T} q^0 f_t(q^0)$. We next divide the analysis depending whether a promotion occurs at time t or not.

Case 1: Time t is not a promotion period, so that t is between two consecutive promotion periods $t_i < t < t_{i+1}$ (or t is after the last promotion). In this case, we have: $POP_t = q^0 \left[f_t(q^0) + e_t^1 + \cdots + e_t^i \right]$. The LP objective at time t is given by:

$$LP_t = q^0 f_t(q^0) + \sum_{j=1}^{i} \left(q^0 \left[f_t(q^0) + e_t^j \right] - q^0 f_t(q^0) \right) = q^0 \left[f_t(q^0) + e_t^1 + \cdots + e_t^i \right]. \quad (4.36)$$

As a result, at each time t without a promotion, we have $POP_t = LP_t$ and hence equation (4.24) is satisfied. We next consider the second case.

Case 2: Time t is a promotion period, i.e., $t = t_i$ for some i. In this case, we obtain:

$$POP_t = q^{k_i}\left[f_{t_i}(q^{k_i}) + e^1_{t_i} + \cdots + e^{i-1}_{t_i}\right]. \tag{4.37}$$

The LP objective at time t is composed of three different parts. First, if $t_j < t$, then the contribution of $POP(\mathbf{p}_{(t_j,k_j)})$ at time t is equal to: $q^0\left[f_t(q^0) + e^j_{t_i}\right]$. Second, if $t_j = t = t_i$, then the contribution of $POP(\mathbf{p}_{(t_j,k_j)})$ at time t is equal to: $q^{k_i} f_{t_i}(q^{k_i})$. Third, if $t_j > t$, then the contribution of $POP(\mathbf{p}_{(t_j,k_j)})$ at time t is the same as the contribution of $POP(\mathbf{p}^0)$ at time t. Therefore, in a similar way as in equation (4.36), the LP objective at time t can be written as:

$$LP_t = \sum_{j=1}^{i-1} q^0 e^j_{t_i} + q^{k_i} f_{t_i}(q^{k_i}). \tag{4.38}$$

By comparing equations (4.37) and (4.38), one can see that equation (4.24) is satisfied and this concludes the proof of the first claim.

The second claim is a consequence of the first one. The first inequality follows from the facts that $q^{k_j} - q^0 \leq 0$ and $g_{t_j - t_i}(q^{k_j}) \leq 0$. The second inequality follows from the facts that $0 \geq q^{k_j} - q^0 \geq q^K - q^0$, and $t_j - t_i \geq (j-i)(S+1)$ (from the constraints on separating periods between successive promotions). By using the properties of the functions g_k from Assumption 12, we obtain: $0 \geq g_{t_j - t_i}(q^{k_j}) \geq g_{(j-i)(S+1)}(q^K)$. □

4.E Proofs of Tightness for Additive Demand

1. Lower bound

Proof. In the case when $S \geq M$, we know from Proposition 4.5.1 that an optimal solution of the LP is also an optimal solution of the POP. Thus, the result holds in this case.

In the case when $S < M$, we will construct a POP problem:

$$POP(\langle q^k\rangle_{k=0}^{K}, \langle f_t\rangle_{t=1}^{T}, \langle c_t\rangle_{t=1}^{T}, \langle g_m\rangle_{m=1}^{M}, L, S),$$

and a price vector $\mathbf{p}^*$, which we will show is both an LP optimal solution and a POP optimal solution. Let $T = L(M+1)$. Let us define the price vector $\mathbf{p}^*$ by:

$$p_t^* = \begin{cases} q^K & t \in \mathbf{U}, \\ q^0 & t \notin \mathbf{U}. \end{cases}$$

Let $\mathbf{U} = \{1, (M+1)+1, 2(M+1)+1, \ldots, (L-1)(M+1)+1\}$ denote the promotion periods of $\mathbf{p}^*$.

Let us define $Y = \sum_{i=1}^{M} |g_i(q^K)|$ and $Z = (L+1)q^0 Y/q^K$ and the demand functions f_t to be:

$$f_t(p_t) = \begin{cases} Z & \text{if } t \in \mathbf{U} \text{ and } p_t = q^K, \\ Y & \text{otherwise.} \end{cases}$$

Note that for any feasible price vector $\mathbf{p}$, the demand at each time is nonnegative. Let us define the costs $c_t = 0, \forall t = 1, \ldots, T$. We prove the proposition by the following steps:

Step 1: We show that an optimal LP solution is the price vector $\mathbf{p}^{LP} = \mathbf{p}^*$.

Step 2: We show that an optimal POP solution is the price vector $\mathbf{p}^{POP} = \mathbf{p}^*$.

Proof of Step 1 By defintion, we have: $POP(\mathbf{p}\{(t,K)\}) - POP(\mathbf{p}^0) = q^K Z - q^0 Y - q^0 Y$ for $t \in \mathbf{U}$. The first term is the period t profit of $POP(\mathbf{p}\{(t,K)\})$, the second term is the period t profit of $POP(\mathbf{p}^0)$, and the third term is the reduction in profit of periods $t+1, \ldots, t+M$ of $POP(\mathbf{p}\{(t,K)\})$ due to the promotion in period t. Therefore, the

LP coefficients as defined in (4.6) are:

$$b_t^k = \begin{cases} q^K Z - 2q^0 Y \geq 0 & \text{if } t \in \mathbf{U}, k = K \\ \leq 0 & \text{otherwise} \end{cases}$$

The LP optimal solution selects at most L of γ_t^k, for $k = 1, \ldots, K$ to be 1. Consequently, the optimal LP objective is bounded above by $Tq^0Y + L(q^K Z - 2q^0 Y)$. In fact, the following γ^* corresponding to $\mathbf{p}^*$ achieves this bound and is therefore optimal:

$$(\gamma^*)_t^k = \begin{cases} 1 & \text{if } t \in \mathbf{U}, k = K \\ 1 & \text{if } t \notin \mathbf{U}, k = 0 \\ 0 & \text{otherwise} \end{cases}$$

We conclude that $\mathbf{p}^{LP} = \mathbf{p}^*$.

Proof of Step 2 We show that for any feasible price vector $\mathbf{p}$, we have $POP(\mathbf{p}^*) \geq POP(\mathbf{p})$. Observe that the POP profit for $\mathbf{p}^*$ is given by:

$$POP(\mathbf{p}^*) = Lq^K Z + (T - L)q^0 Y - Lq^0 Y.$$

In particular, the first term corresponds to the profit from the promotion periods $\mathbf{U}$ and the second term is the profit from the non-promotion periods $\mathbb{T} \setminus \mathbf{U}$ before promotions. Finally, the third term represents the reduction in profit during the non-promotion periods due to the promotions in $\mathbf{U}$.

Let POP_t be the $POP(\mathbf{p})$ profit at period t. If we promote at time $t \in \mathbf{U}$ using the price q^K, then $POP_t = q^K Z$ and otherwise, $POP_t \leq q^0 Y$. For any $\mathbf{p} \neq \mathbf{p}^*$, $\mathbf{p}$ has at most $L - 1$ promotions at the time periods $t \in \mathbf{U}$. Therefore, we obtain: $POP(\mathbf{p}) \leq (L-1)q^K Z + (T - L + 1)q^0 Y$. The first term results from the promotions during the periods in $\mathbf{U}$, whereas the second term comes from the non-promotion periods. One can see that:

$$POP(\mathbf{p}^*) - POP(\mathbf{p}) = Lq^K Z + (T-L)q^0 Y - Lq^0 Y - \left[(L-1)q^K Z + (T-L+1)q^0 Y\right]$$
$$= q^K Z - (L+1)q^0 Y \geq 0,$$

from the definition of Z. Therefore, $POP(\mathbf{p}^*) \geq POP(\mathbf{p})$ as desired. □

2. **Upper bound**

Proof. In the case when $S \geq M$, we know from Proposition 4.5.1 that an optimal solution of the LP is also an optimal solution of the POP. We also know from equation (4.22) that $\overline{R} = 0$. Thus, the result holds in this case.

In the case when $S < M$, we will construct a POP problem:

$$POP(\langle q^k \rangle_{k=0}^K, \langle f_t \rangle_{t=1}^T, \langle c_t \rangle_{t=1}^T, \langle g_m \rangle_{m=1}^M, L, S),$$

an optimal LP price vector $\mathbf{p}^{LP}$, and an optimal POP price vector $\mathbf{p}^{POP}$, such that $POP(\mathbf{p}^{POP}) = POP(\mathbf{p}^{LP}) + \overline{R}$. Let $T = (M+1)L$. Let us define $Y = \sum_{i=1}^M |g_i(q^K)|$, $Z = (L+1)q^0 Y / q^K$ and the demand functions f_t to be:

$$f_t(p_t) = \begin{cases} Z & \text{if } 1 \leq t \leq LM+1 \text{ and } p_t = q^K, \\ Y & \text{otherwise.} \end{cases}$$

Note that for any feasible price vector $\mathbf{p}$, the demand at each time is nonnegative. We prove the proposition by the following steps:

Step 1: We show that the following price vector is an optimal LP solution:

$$\mathbf{p}^{LP} = \Big(q^K, \underbrace{q^0, \ldots, q^0}_{M \text{ times}}, q^K, \underbrace{q^0, \ldots, q^0}_{M \text{ times}}, \ldots, q^K, \underbrace{q^0, \ldots, q^0}_{M \text{ times}}\Big).$$

Step 2: We show that the following price vector is an optimal POP solution:

$$\mathbf{p}^{POP} = \Big(q^K, \underbrace{q^0, \ldots, q^0}_{S \text{ times}}, q^K, \underbrace{q^0, \ldots, q^0}_{S \text{ times}}, \ldots, q^K, \underbrace{q^0, \ldots, q^0}_{T-(L-1)(S+1)-1 \text{ times}}\Big).$$

Step 3: We show that $POP(\mathbf{p}^{POP}) = POP(\mathbf{p}^{LP}) + \overline{R}$ which concludes the proof.

Proof of Step 1. By definition, we have:

$$POP(\mathbf{p}\{(t,K)\}) - POP(\mathbf{p}^0) = q^K Z - q^0 Y - q^0 Y$$

for $t \in \mathbf{U}$. The first term is the period t profit of $POP(\mathbf{p}\{(t,K)\})$, the second term is the period t profit of $POP(\mathbf{p}^0)$, and the third term is the reduction in profit of periods $t+1, \ldots, t+M$ of $POP(\mathbf{p}\{(t,K)\})$ due to the promotion in period t. Therefore, the LP coefficients as defined in (4.6) are:

$$b_t^k = \begin{cases} q^K Z - 2q^0 Y \geq 0 & \text{if } t \in \mathbf{U}, k = K \\ \leq 0 & \text{otherwise} \end{cases}$$

The LP optimal solution selects at most L of γ_t^k, for $k = 1, \ldots, K$ to be 1. Consequently, the optimal LP objective is bounded above by $Tq^0Y + L(q^K Z - 2q^0 Y)$. In fact, the following γ^{LP} corresponding to $\mathbf{p}^{LP}$ achieves this bound and is therefore optimal:

$$(\gamma^{LP})_t^k = \begin{cases} 1 & \text{if } t \in \mathbf{U}, k = K \\ 1 & \text{if } t \notin \mathbf{U}, k = 0 \\ 0 & \text{otherwise} \end{cases}$$

We conclude that $\mathbf{p}^{LP}$ is an optimal solution to LP. Note that because any two promotions are separated by at least M periods, $ER(\mathbf{p}^{LP}) = 0$ and then from Proposition

4.6.6:

$$POP(\mathbf{p}^{LP}) = LP(\mathbf{p}^{LP}). \tag{4.39}$$

Proof of Step 2. By using Proposition 4.6.6, we know that for any feasible price vector $\mathbf{p}$: $POP(\mathbf{p}) = LP(\mathbf{p}) + ER(\mathbf{p})$. One can see that $LP(\mathbf{p}) \leq LP(\mathbf{p}^{POP})$. Indeed, we note that the price vector $\mathbf{p}^{POP}$ is also optimal for the LP by using a similar argument as for $\mathbf{p}^{LP}$. In other words, in this case, both $\mathbf{p}^{LP}$ and $\mathbf{p}^{POP}$ are optimal LP solutions. By using the definition of $\overline{R}$ from (4.22), one can see that $ER(\mathbf{p}) \leq \overline{R}$ for all feasible $\mathbf{p}$. In other words, $\overline{R}$ corresponds to the largest possible error term. In addition, we have in this case: $ER(\mathbf{p}^{POP}) = \overline{R}$ by construction. Since $LP(\mathbf{p}) \leq LP(\mathbf{p}^{POP})$ and $ER(\mathbf{p}) \leq ER(\mathbf{p}^{POP})$ for any $\mathbf{p}$, we obtain $POP(\mathbf{p}) \leq POP(\mathbf{p}^{POP})$ for any $\mathbf{p}$ so that $\mathbf{p}^{POP}$ is an optimal POP solution. In addition, we have shown that:

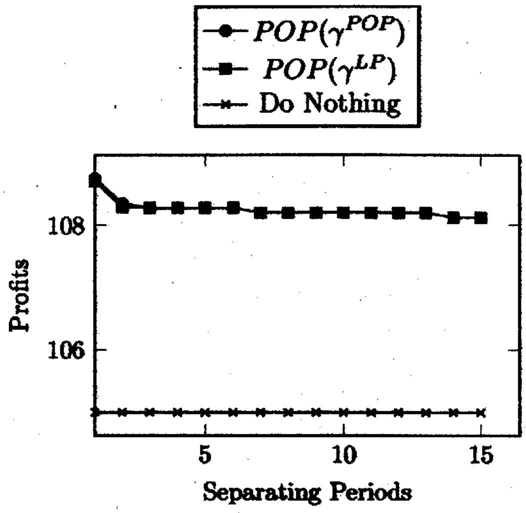
$$POP(\mathbf{p}^{POP}) = LP(\mathbf{p}^{POP}) + \overline{R}. \tag{4.40}$$

Proof of Step 3. In the proof of Step 2 we have shown that $LP(\mathbf{p}^{LP}) = LP(\mathbf{p}^{POP})$. Combining this equation with (4.39) and (4.40) gives us the desired result. □

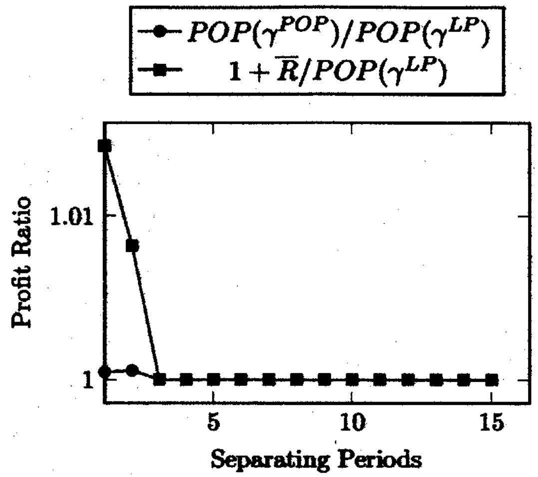
4.F Additive demand: illustrating the bounds

In what follows, we test numerically the upper bound for the additive model from Section 4.6.2 by varying the different model parameters. In Figures 4.F.1, 4.F.2 and 4.F.3, the demand model is given by: $d_t(\mathbf{p}) = 30 - 50p_t + 15p_{t-1} + 10p_{t-2} + 5p_{t-3}$.

Dependence on separating periods: In Figure 4.F.1, we vary the number of separating periods S. We make the following observations: *a*) As one would expect from Proposition 4.5.1, the LP approximation coincides with the optimal POP solution for $S \geq M = 4$. *b*) As expected, as S increases, the upper bound $1 + \overline{R}/POP(\gamma^{LP})$ decreases. Indeed, the larger is S, the more separated promotions are and as a result, it



(a) Profits



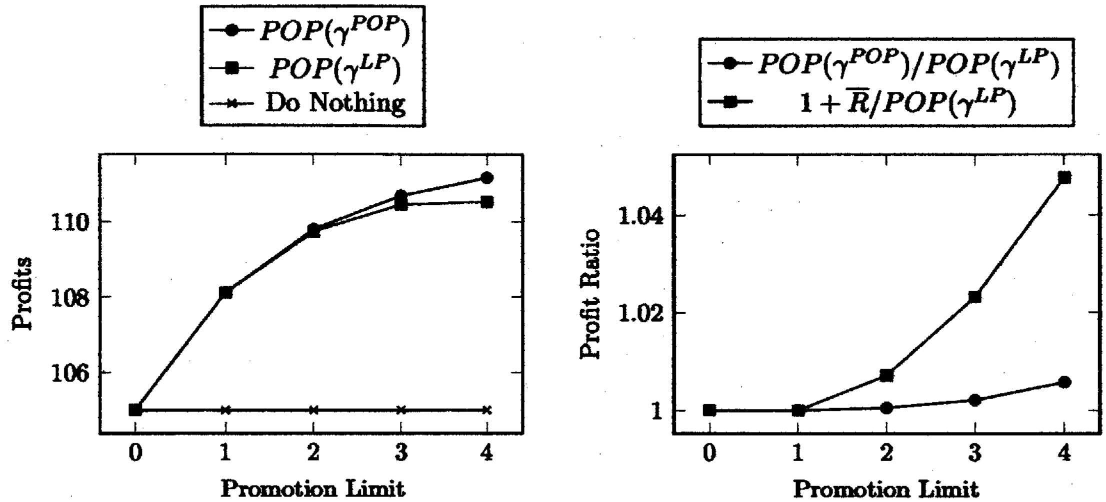
(b) Profit ratio

Figure 4.F.1: Results of Additive Demand Model (Varying Separation)

Note. Example parameters: $L = 3, Q = \{1, 0.95, 0.90, 0.85, 0.80, 0.75, 0.70\}$.

reduces the interaction between promotions which are neglected in the LP approximation. *c*) For any value of S, the upper bound on the relative optimality gap (between the POP objective at optimality versus evaluated at the LP approximation solution) is at most 2.5%, whereas the realized one is less than 1.5%. In practice, typically the number of separating periods is at least 2.

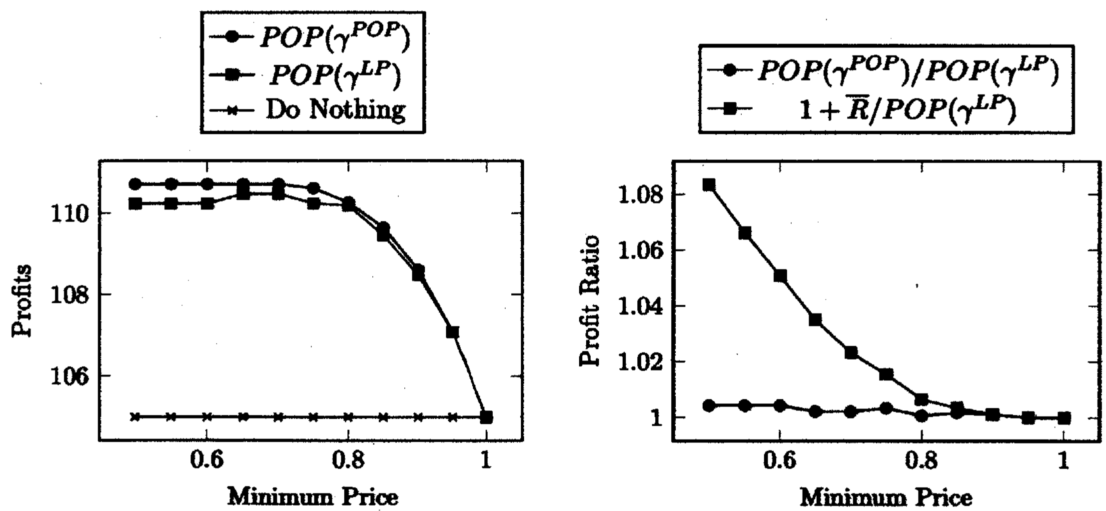
Dependence on the number of promotions allowed: In Figure 4.F.2, we vary the number of promotions allowed L. We make the following observations: *a*) As one would expect from Proposition 4.5.1, the LP approximation coincides with the optimal POP solution for $L = 1$. *b*) For $L \leq 6$ (recall that $T = 13$), the upper bound on the relative optimality gap is at most 10%. As expected, the upper bound increases as L increases. This follows from the definition of $\overline{R}$ in Theorem 4.6.5. Unlike the multiplicative case for which $\underline{R}$ was asymptotically converging as L increases; in the additive case, $\overline{R}$ can grow to infinity as L increases.



(a) Profits (b) Profit ratio

Figure 4.F.2: Results of Additive Demand Model (Varying Promotion Limit)

Note. Example parameters: $S = 0, Q = \{1, 0.95, 0.90, 0.85, 0.80, 0.75, 0.70\}$.

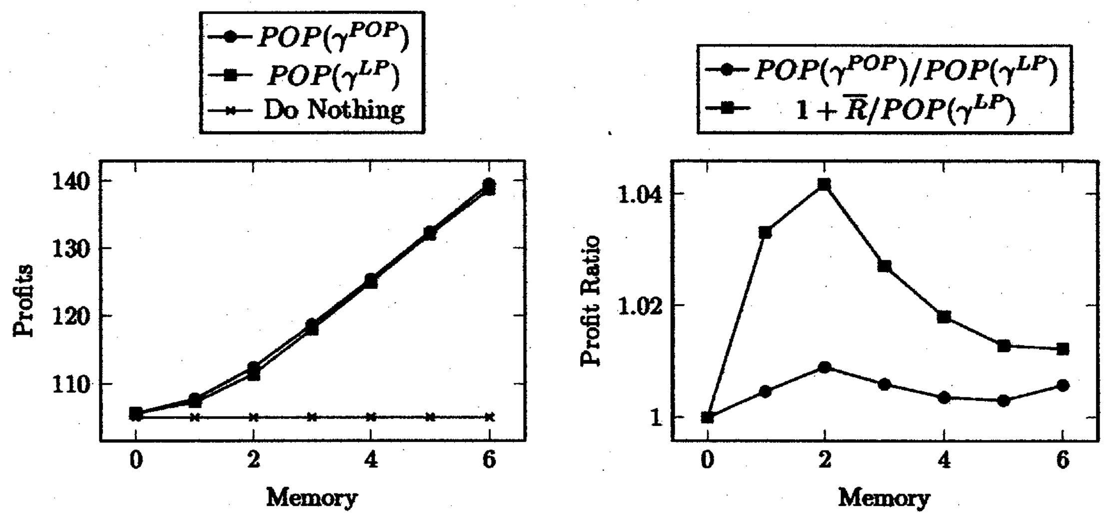


(a) Profits (b) Profit ratio

Figure 4.F.3: Results of Additive Demand Model (Varying Minimum Price)

Note. Example parameters: $L = 3, S = 0$.

Dependence on the minimal price of the price ladder: In Figure 4.F.3, we vary the minimum promotion price q^K. We make the following observations: *a*) As one would expect, the LP approximation coincides with the optimal POP solution for $q^K = 1$, i.e., the promotion price is equal to the regular price at all times. *b*) The upper bound on the relative optimality gap is at most 2.5%. From the definition of $\overline{R}$ in Theorem 4.6.5, one can see that the additive contribution $\overline{R}$ increases as q^K decreases.



(a) Profits (b) Profit ratio

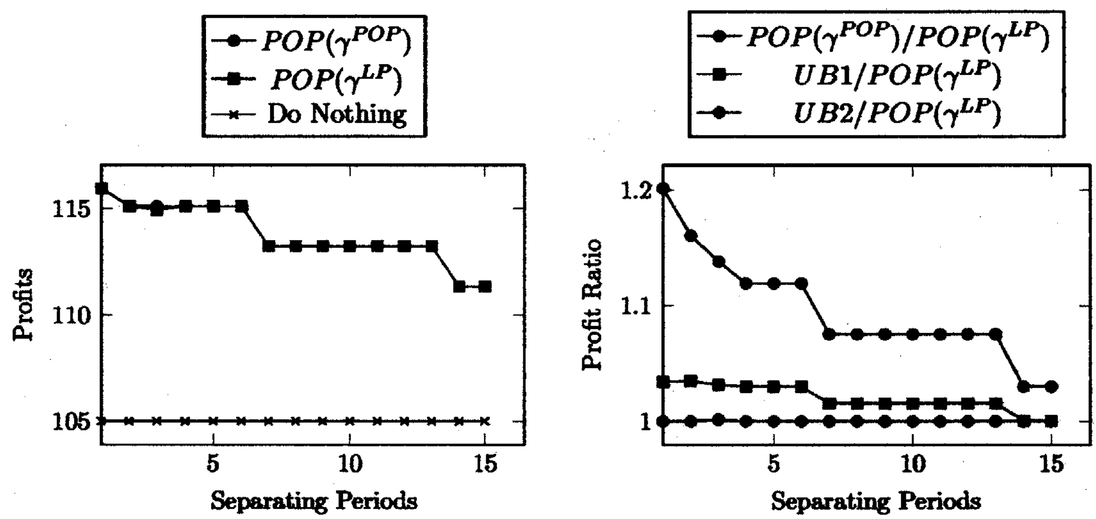
Figure 4.F.4: Results of Additive Demand Model (Varying Memory)

Note. Example parameters: $d_t(\mathbf{p}) = 30-(20+20M)p_t+20p_{t-1}+20p_{t-2}+\cdots+20p_{t-M}$, $L = 3, S = 0$.

Dependence on the length of the memory: In Figure 4.F.4, we vary the memory parameter M. Note that in this example, we have chosen equal coefficients for $g_1, g_2, \ldots, g_M$, as a "worst case" so that past prices have a uniformly strong effect on current demand. We make the following observations: *a*) As one would expect from Proposition 4.5.1, the LP approximation coincides with the optimal POP solution for $S \geq M$, i.e., $M = 0$. *b*) The upper bound on the relative optimality gap is at most 4.5%. From the definition of $\overline{R}$ in Theorem4.6.5, one can see that $\overline{R}$ increases with M, until it hits the constraint on the limited number of promotions (in this case is $L = 3$). In particular, we have two cases. When $M < L$, increasing the memory parameter by one unit will increase $\overline{R}$. Indeed, from the definition of $\overline{R}$, some of the terms $g_{(j-i)(S+1)}(q^K)$ will switch from zero to a negative value. When $M > L$, increasing the memory parameter by one one will not increase $\overline{R}$. In this case, the terms $g_{(j-i)(S+1)}(q^K)$ do not change.

4.G Unified demand: illustrating the bounds

In what follows, we test numerically the upper bounds for the unified model from Section 4.6.3 by varying the different parameters of the model. In Figures 4.G.1, 4.G.2 and 4.G.3, the demand model is given by: $d_t(\mathbf{p}) = 0.5d^{\text{mult}}(\mathbf{p}) + 0.5d^{\text{add}}(\mathbf{p})$, where: $d^{\text{mult}}(\mathbf{p}) = 10p_t^{-4}p_{t-1}^{0.5}p_{t-2}^{0.3}p_{t-3}^{0.2}p_{t-4}^{0.1}$, $d^{\text{add}}(\mathbf{p}) = 30 - 50p_t + 15p_{t-1} + 10p_{t-2} + 5p_{t-3}$. We next illustrate both upper bounds $UB1$ and $UB2$ from equations (4.31) and (4.30) respectively as well as the performance of the LP approximation for the above unified demand model.



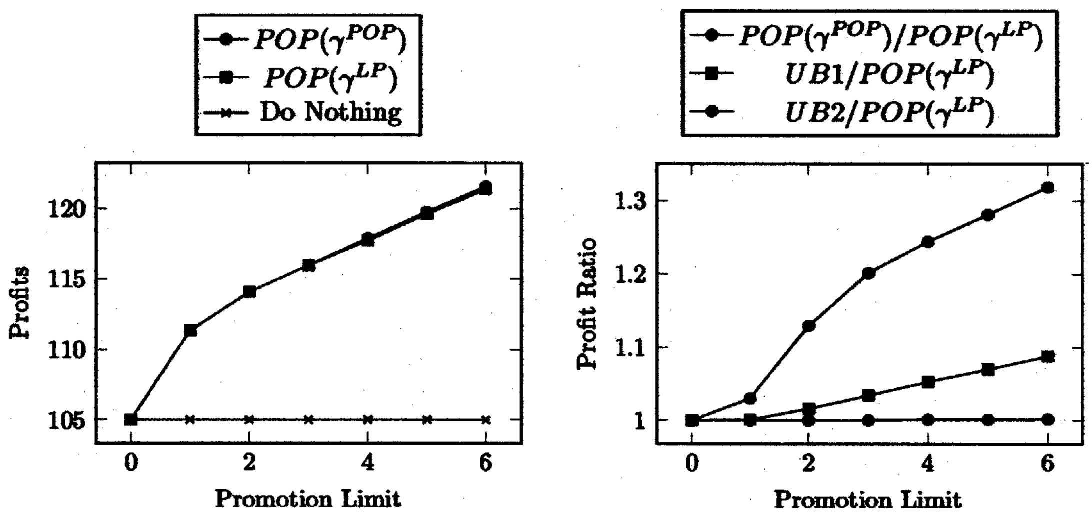
(a) Profits

(b) Profit ratio

Figure 4.G.1: Results of Unified Demand Model (Varying Separation)

Note. Example parameters: $L = 3, Q = \{1, 0.95, 0.90, 0.85, 0.80, 0.75, 0.70\}$.

Dependence on separating periods: In Figure 4.G.1, we vary the number of separating periods S. We make the following observations: *a*) As one would expect from Proposition 4.5.1, the LP approximation coincides with the optimal POP solution for $S \geq M = 4$. *b*) As expected, as S increases, the upper bound $1 + \overline{R}/POP(\gamma^{LP})$ decreases. Indeed, the larger is S, the more separated promotions are and hence it reduces the interaction between promotions neglected in the LP approximation. *c*) For any value of S, the upper bound $UB1$ on the relative optimality gap (between the POP objective at optimality versus evaluated at the LP approximation solution) is less than 4%. In practice, typically the number of separating periods is at least 2.



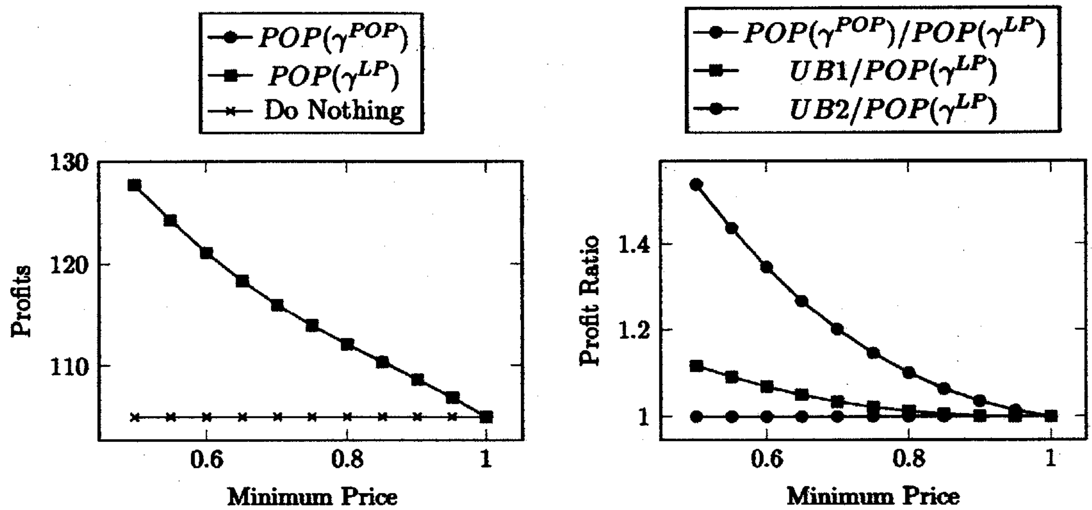
(a) Profits

(b) Profit ratio

Figure 4.G.2: Results of Unified Demand Model (Varying Promotion Limit)

Note. Example parameters: $S = 1, Q = \{1, 0.95, 0.90, 0.85, 0.80, 0.75, 0.70\}$.

Dependence on the number of promotions allowed: In Figure 4.G.2, we vary the number of promotions allowed L. We make the following observations: *a*) As one would expect from Proposition 4.5.1, the LP approximation coincides with the optimal POP solution for $L = 1$. *b*) For $L \leq 6$ (recall that $T = 13$), the upper bound $UB1$ on the relative optimality gap is at most 9%. However, the upper bound will continue to grow with L. This follows from the additive part of the demand for which $\overline{R}$ in (4.22) is increasing with L. Unlike the multiplicative case for which $\underline{R}$ was asymptotically converging as L increases; in the additive case, $\overline{R}$ can grow to infinity as L increases. Consequently, for any unified model with $0 \leq \lambda < 1$, the additive upper bound contribution will grow with respect to L.



(a) Profits

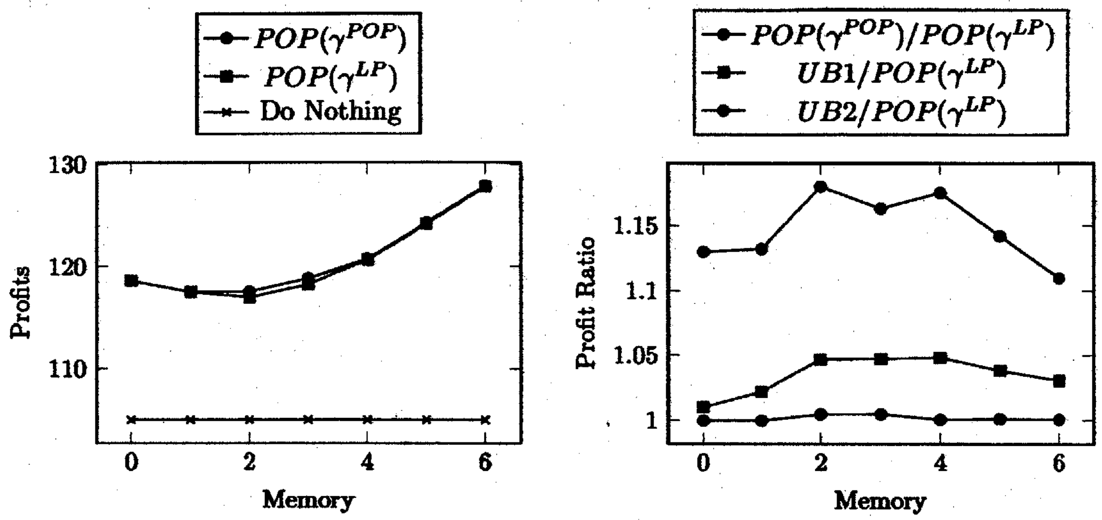
(b) Profit ratio

Figure 4.G.3: Results of Unified Demand Model (Varying Minimum Price)

Note. Example parameters: $L = 3, S = 1$.

Dependence on the minimal price of the price ladder: In Figure 4.G.3, we vary the minimum promotion price q^K. We make the following observations: *a*) As one would expect, the LP approximation coincides with the optimal POP solution when $q^K = 1$, i.e., the promotion price is equal to the regular price at all times. *b*) The upper bound $UB1$ on the relative optimality gap is at most 11%. From the definition of $\overline{R}$ in Theorem 4.6.5, one can see that $\overline{R}$ increases as q^K decreases.

Dependence on the length of the memory: In Figure 4.G.4, we vary the memory parameter M. Note that in this example, we have chosen equal coefficients for $g_1, g_2, \ldots, g_M$, as a "worst case" so that past prices also have a uniformly strong effect on current demand. We make the following observations: *a*) As one would expect from



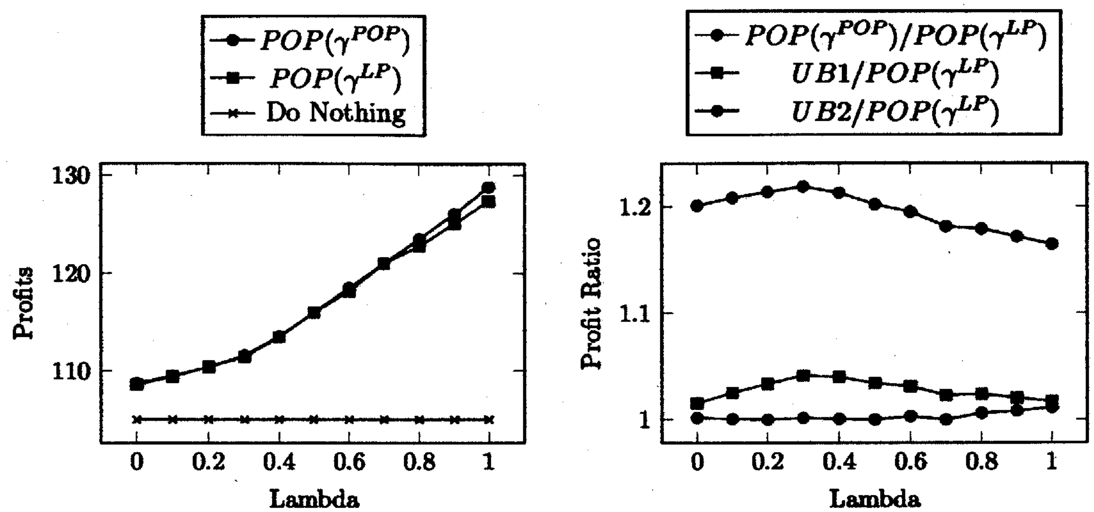
(a) Profits (b) Profit ratio

Figure 4.G.4: Results of Unified Demand Model (Varying Memory)

Note. Example parameters: $d_t(\mathbf{p}) = 0.5(10p_t^{-4}p_{t-1}^{0.2}p_{t-2}^{0.2}\cdots p_{t-M}^{0.2}) + 0.5[30 - (20 + 20M)p_t + 20p_{t-1} + 20p_{t-2} + \cdots + 20p_{t-M}]$, $L = 3, S = 1$.

Proposition 4.5.1, the LP approximation coincides with the optimal POP solution for $S \geq M$, i.e., $M = 0$. *b)* The upper bound $UB1$ on the relative optimality gap is at most 5%. From the definition of $\overline{R}$ in (4.22), one can see that the additive contribution $\overline{R}$ increases with M, until it hits the constraint on the limited number of promotions (in this case is $L = 3$). In particular, we have two cases. When $M < L$, increasing the memory parameter by one unit will increase $\overline{R}$. Indeed, from the definition of $\overline{R}$, some of the terms $g_{(j-i)(S+1)}(q^K)$ will switch from zero to a negative value. When $M > L$, increasing the memory parameter by one one will not increase $\overline{R}$. In this case, the terms $g_{(j-i)(S+1)}(q^K)$ do not change.

Dependence on the the parameter λ: Note that when λ is set to either 0 or 1, we retrieve the bounds for the additive and multiplicative models respectively. As one can see from Figure 4.G.5, the upper bound, $UB1$ is better than the second bound $UB2$ for any value of $0 \leq \lambda \leq 1$. In addition, the upper bound $UB1$ achieves its worst value of 4.5% when $\lambda = 0.3$ for which the model is a mixture of both demand forms. In other



(a) Profits (b) Profit ratio

Figure 4.G.5: Results of Unified Demand Model (Varying λ)

Note. Example parameters: $d^{\text{mult}}(p) = 10p_t^{-4}p_{t-1}^{0.2}\cdots p_{t-M}^{0.2}$, $d^{\text{add}}(p) = 30 - 20p_t + 2p_{t-1} + \cdots + 2p_{t-M}$, $d_t(\mathrm{p}) = \lambda d^{\text{mult}}(p) + (1-\lambda)d^{\text{add}}(p)$, $L = 3, S = 1$.

words, the bound is better when computed for each segment separately but achieves its worst case for some given combination of both segments (in this case, $\lambda = 0.3$).

Bibliography

Kusum L Ailawadi, Karen Gedenk, Christian Lutzky, and Scott A Neslin. Decomposition of the sales impact of promotion-induced stockpiling. *Journal of Marketing Research*, pages 450–467, 2007.

D. Bertsimas and R. Weismantel. *Optimization over integers*, volume 13. Dynamic Ideas Belmont, 2005.

Dimitris Bertsimas and Romy Shioda. Algorithm for cardinality-constrained quadratic optimization. *Computational Optimization and Applications*, 43(1):1–22, 2009.

Daniel Bienstock. Computational study of a family of mixed-integer quadratic programming problems. *Mathematical programming*, 74(2):121–140, 1996.

R. C. Blattberg and S. A. Neslin. *Sales promotion: Concepts, Methods, and Strategies.* Prentice Hall, 1990.

Niv Buchbinder, Moran Feldman, Joseph Seffi Naor, and Roy Schwartz. A tight linear time (1/2)-approximation for unconstrained submodular maximization. In *Foundations of Computer Science (FOCS), 2012 IEEE 53rd Annual Symposium on*, pages 649–658. IEEE, 2012.

Gérard P Cachon and Martin A Lariviere. Supply chain coordination with revenue-sharing contracts: strengths and limitations. *Management science*, 51(1):30–44, 2005.

Katia Campo, Els Gijsbrechts, and Patricia Nisol. Towards understanding consumer response to stock-outs. *Journal of Retailing*, 76(2):219–242, 2000.

Felipe Caro and Jérémie Gallien. Clearance pricing optimization for a fast-fashion retailer. *Operations Research*, 60(6):1404–1422, 2012.

Xin Chen, Peng Hu, Stephen Shum, and Yuhan Zhang. Stochastic inventory model with reference price effects. 2013. Submitted.

L. G. Cooper, P. Baron, W. Levy, M. Swisher, and P. Gogos. Promocast: A new forecasting method for promotion planning. *Marketing Science*, 18(3):301–316, 1999.

Daniel Corsten and Thomas Gruen. Stock-outs cause walkouts. *Harvard Business Review*, 82(5):26–28, 2004.

Uriel Feige, Vahab S Mirrokni, and Jan Vondrak. Maximizing non-monotone submodular functions. *SIAM Journal on Computing*, 40(4):1133–1153, 2011.

G. Fibich, A. Gavious, and O. Lowengart. Explicit solutions of optimization models and differential games with nonsmooth (asymmetric) reference-price effects. *Operations Research*, 51(5):721–734, 2003.

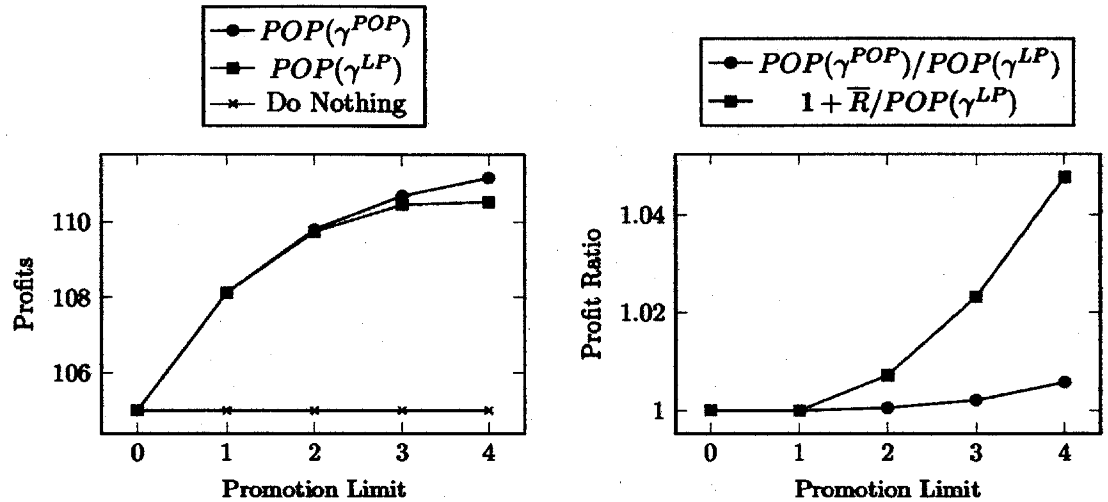
Roger Fletcher and Sven Leyffer. Solving mixed integer nonlinear programs by outer approximation. *Mathematical programming*, 66(1-3):327–349, 1994.

Eijte W Foekens, Peter SH Leeflang, and Dick R Wittink. Varying parameter models to accommodate dynamic promotion effects. *Journal of Econometrics*, 89(1):249–268, 1998.

Ignacio E Grossmann. Review of nonlinear mixed-integer and disjunctive programming techniques. *Optimization and Engineering*, 3(3):227–252, 2002.

Sunil Gupta. Impact of sales promotions on when, what, and how much to buy. *Journal of Marketing Research*, pages 342–355, 1988.

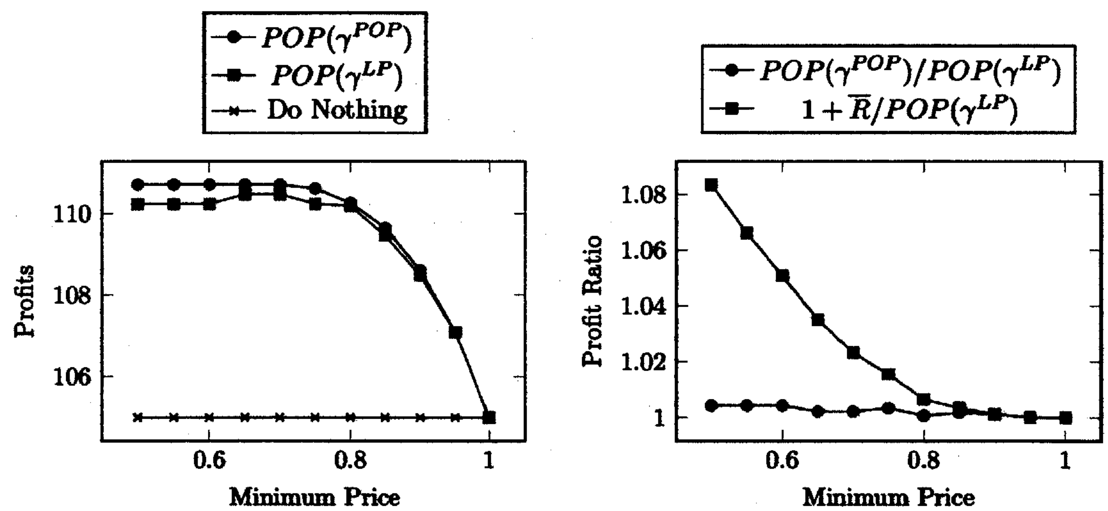
H. J. Van Heerde, P. S. H. Leeflang, and D. R. Wittink. The estimation of pre- and post- promotion dips with store-level scanner data. *Journal of Marketing Research*, 37:383–395, 2000.



(a) Profits (b) Profit ratio

Figure 4.F.2: Results of Additive Demand Model (Varying Promotion Limit)

Note. Example parameters: $S = 0, \mathcal{Q} = \{1, 0.95, 0.90, 0.85, 0.80, 0.75, 0.70\}$.

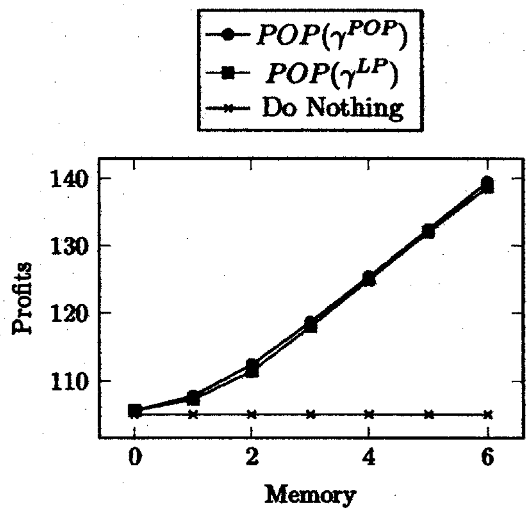


(a) Profits (b) Profit ratio

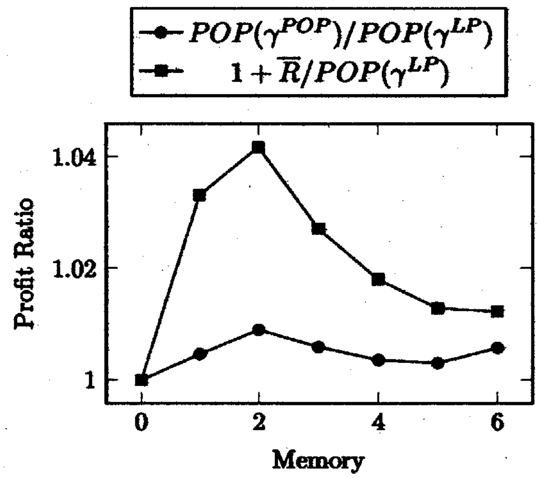
Figure 4.F.3: Results of Additive Demand Model (Varying Minimum Price)

Note. Example parameters: $L = 3, S = 0$.

Dependence on the minimal price of the price ladder: In Figure 4.F.3, we vary the minimum promotion price q^K. We make the following observations: *a*) As one would expect, the LP approximation coincides with the optimal POP solution for $q^K = 1$, i.e., the promotion price is equal to the regular price at all times. *b*) The upper bound on the relative optimality gap is at most 2.5%. From the definition of $\overline{R}$ in Theorem 4.6.5, one can see that the additive contribution $\overline{R}$ increases as q^K decreases.



(a) Profits



(b) Profit ratio

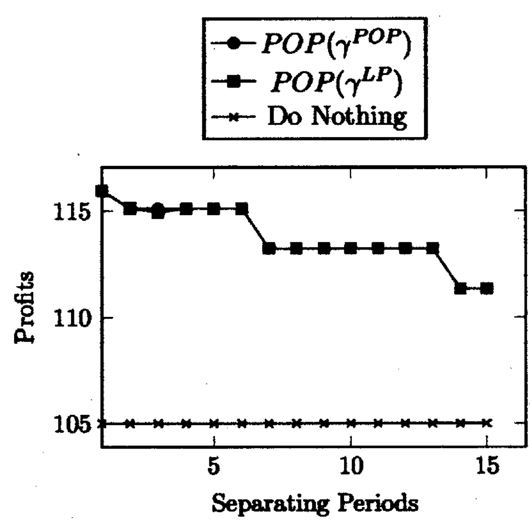
Figure 4.F.4: Results of Additive Demand Model (Varying Memory)

Note. Example parameters: $d_t(\mathbf{p}) = 30-(20+20M)p_t+20p_{t-1}+20p_{t-2}+\cdots+20p_{t-M}$, $L = 3, S = 0$.

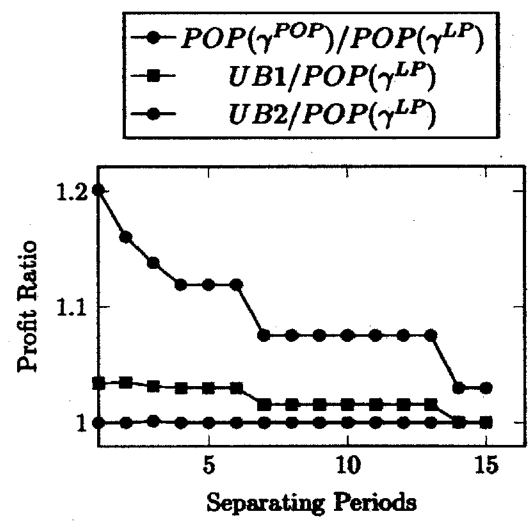
Dependence on the length of the memory: In Figure 4.F.4, we vary the memory parameter M. Note that in this example, we have chosen equal coefficients for $g_1, g_2, \ldots, g_M$, as a "worst case" so that past prices have a uniformly strong effect on current demand. We make the following observations: *a*) As one would expect from Proposition 4.5.1, the LP approximation coincides with the optimal POP solution for $S \geq M$, i.e., $M = 0$. *b*) The upper bound on the relative optimality gap is at most 4.5%. From the definition of $\overline{R}$ in Theorem4.6.5, one can see that $\overline{R}$ increases with M, until it hits the constraint on the limited number of promotions (in this case is $L = 3$). In particular, we have two cases. When $M < L$, increasing the memory parameter by one unit will increase $\overline{R}$. Indeed, from the definition of $\overline{R}$, some of the terms $g_{(j-i)(S+1)}(q^K)$ will switch from zero to a negative value. When $M > L$, increasing the memory parameter by one one will not increase $\overline{R}$. In this case, the terms $g_{(j-i)(S+1)}(q^K)$ do not change.

4.G Unified demand: illustrating the bounds

In what follows, we test numerically the upper bounds for the unified model from Section 4.6.3 by varying the different parameters of the model. In Figures 4.G.1, 4.G.2 and 4.G.3, the demand model is given by: $d_t(\mathbf{p}) = 0.5d^{\text{mult}}(\mathbf{p}) + 0.5d^{\text{add}}(\mathbf{p})$, where: $d^{\text{mult}}(\mathbf{p}) = 10p_t^{-4}p_{t-1}^{0.5}p_{t-2}^{0.3}p_{t-3}^{0.2}p_{t-4}^{0.1}$, $d^{\text{add}}(\mathbf{p}) = 30 - 50p_t + 15p_{t-1} + 10p_{t-2} + 5p_{t-3}$. We next illustrate both upper bounds $UB1$ and $UB2$ from equations (4.31) and (4.30) respectively as well as the performance of the LP approximation for the above unified demand model.



(a) Profits

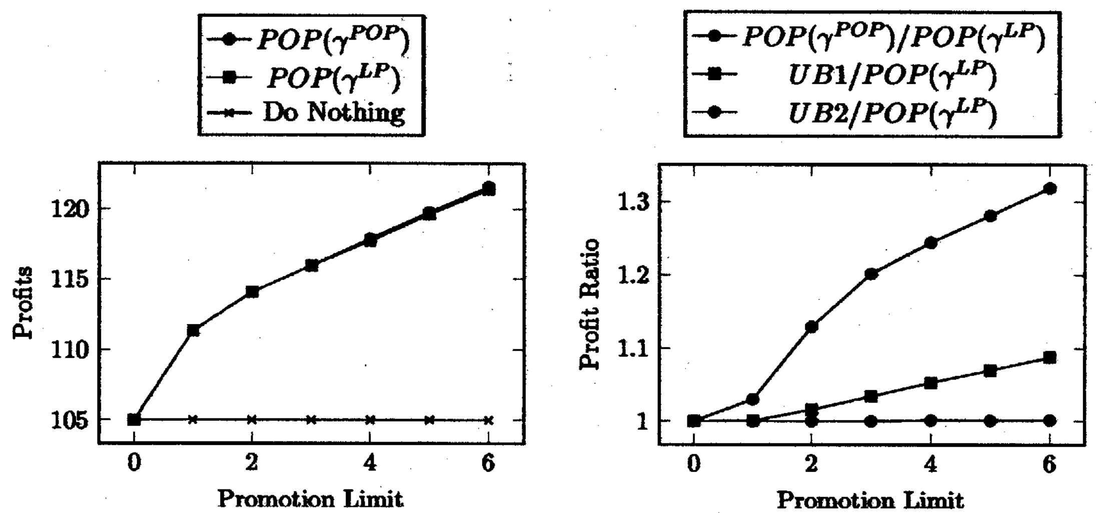


(b) Profit ratio

Figure 4.G.1: Results of Unified Demand Model (Varying Separation)

Note. Example parameters: $L = 3, \mathcal{Q} = \{1, 0.95, 0.90, 0.85, 0.80, 0.75, 0.70\}$.

Dependence on separating periods: In Figure 4.G.1, we vary the number of separating periods S. We make the following observations: *a*) As one would expect from Proposition 4.5.1, the LP approximation coincides with the optimal POP solution for $S \geq M = 4$. *b*) As expected, as S increases, the upper bound $1 + \overline{R}/POP(\gamma^{LP})$ decreases. Indeed, the larger is S, the more separated promotions are and hence it reduces the interaction between promotions neglected in the LP approximation. *c*) For any value of S, the upper bound $UB1$ on the relative optimality gap (between the POP objective at optimality versus evaluated at the LP approximation solution) is less than 4%. In practice, typically the number of separating periods is at least 2.



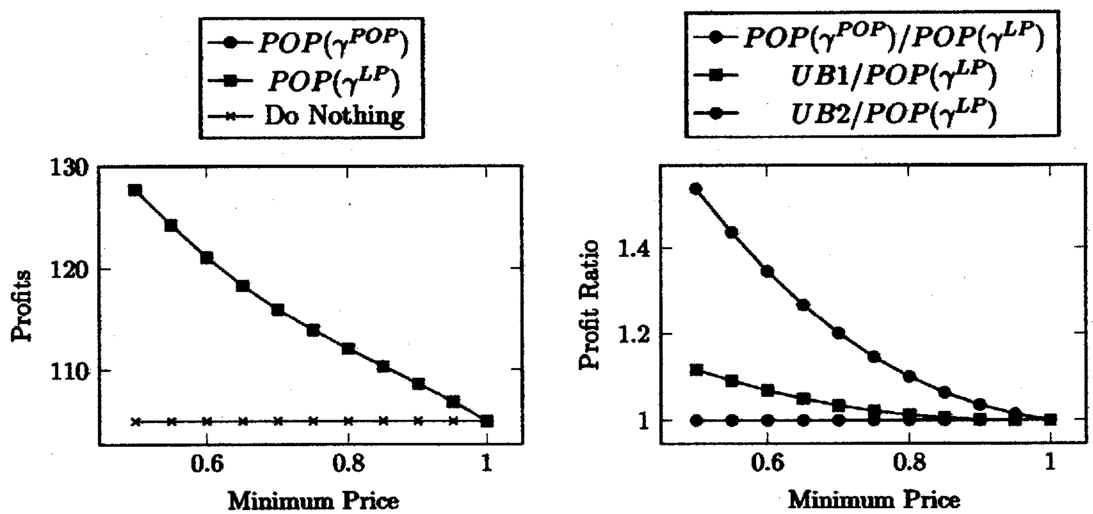
(a) Profits

(b) Profit ratio

Figure 4.G.2: Results of Unified Demand Model (Varying Promotion Limit)

Note. Example parameters: $S = 1, Q = \{1, 0.95, 0.90, 0.85, 0.80, 0.75, 0.70\}$.

Dependence on the number of promotions allowed: In Figure 4.G.2, we vary the number of promotions allowed L. We make the following observations: *a*) As one would expect from Proposition 4.5.1, the LP approximation coincides with the optimal POP solution for $L = 1$. *b*) For $L \leq 6$ (recall that $T = 13$), the upper bound $UB1$ on the relative optimality gap is at most 9%. However, the upper bound will continue to grow with L. This follows from the additive part of the demand for which $\overline{R}$ in (4.22) is increasing with L. Unlike the multiplicative case for which $\underline{R}$ was asymptotically converging as L increases; in the additive case, $\overline{R}$ can grow to infinity as L increases. Consequently, for any unified model with $0 \leq \lambda < 1$, the additive upper bound contribution will grow with respect to L.



(a) Profits

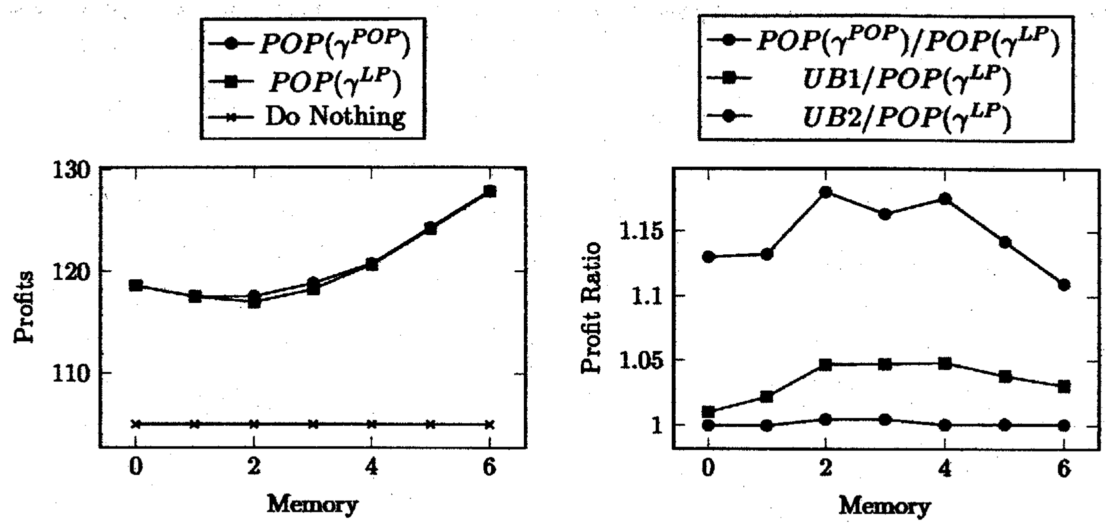
(b) Profit ratio

Figure 4.G.3: Results of Unified Demand Model (Varying Minimum Price)

Note. Example parameters: $L = 3, S = 1$.

Dependence on the minimal price of the price ladder: In Figure 4.G.3, we vary the minimum promotion price q^K. We make the following observations: *a*) As one would expect, the LP approximation coincides with the optimal POP solution when $q^K = 1$, i.e., the promotion price is equal to the regular price at all times. *b*) The upper bound $UB1$ on the relative optimality gap is at most 11%. From the definition of $\overline{R}$ in Theorem 4.6.5, one can see that $\overline{R}$ increases as q^K decreases.

Dependence on the length of the memory: In Figure 4.G.4, we vary the memory parameter M. Note that in this example, we have chosen equal coefficients for $g_1, g_2, \ldots, g_M$, as a "worst case" so that past prices also have a uniformly strong effect on current demand. We make the following observations: *a*) As one would expect from



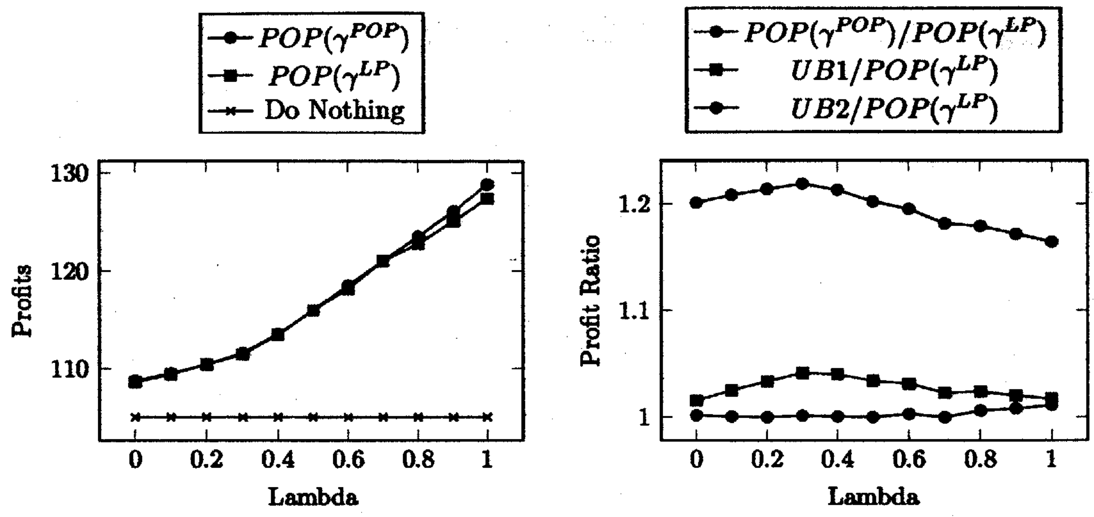
(a) Profits (b) Profit ratio

Figure 4.G.4: Results of Unified Demand Model (Varying Memory)

Note. Example parameters: $d_t(\mathbf{p}) = 0.5\left(10p_t^{-4}p_{t-1}^{0.2}p_{t-2}^{0.2}\cdots p_{t-M}^{0.2}\right) + 0.5\left[30 - (20 + 20M)p_t + 20p_{t-1} + 20p_{t-2} + \cdots + 20p_{t-M}\right]$, $L = 3, S = 1$.

Proposition 4.5.1, the LP approximation coincides with the optimal POP solution for $S \geq M$, i.e., $M = 0$. *b)* The upper bound $UB1$ on the relative optimality gap is at most 5%. From the definition of $\overline{R}$ in (4.22), one can see that the additive contribution $\overline{R}$ increases with M, until it hits the constraint on the limited number of promotions (in this case is $L = 3$). In particular, we have two cases. When $M < L$, increasing the memory parameter by one unit will increase $\overline{R}$. Indeed, from the definition of $\overline{R}$, some of the terms $g_{(j-i)(S+1)}(q^K)$ will switch from zero to a negative value. When $M > L$, increasing the memory parameter by one one will not increase $\overline{R}$. In this case, the terms $g_{(j-i)(S+1)}(q^K)$ do not change.

Dependence on the the parameter λ: Note that when λ is set to either 0 or 1, we retrieve the bounds for the additive and multiplicative models respectively. As one can see from Figure 4.G.5, the upper bound, $UB1$ is better than the second bound $UB2$ for any value of $0 \leq \lambda \leq 1$. In addition, the upper bound $UB1$ achieves its worst value of 4.5% when $\lambda = 0.3$ for which the model is a mixture of both demand forms. In other



(a) Profits

(b) Profit ratio

Figure 4.G.5: Results of Unified Demand Model (Varying λ)

Note. Example parameters: $d^{\text{mult}}(p) = 10p_t^{-4}p_{t-1}^{0.2}\cdots p_{t-M}^{0.2}$, $d^{\text{add}}(p) = 30 - 20p_t + 2p_{t-1} + \cdots + 2p_{t-M}$, $d_t(\mathbf{p}) = \lambda d^{\text{mult}}(p) + (1-\lambda)d^{\text{add}}(p)$, $L = 3, S = 1$.

words, the bound is better when computed for each segment separately but achieves its worst case for some given combination of both segments (in this case, $\lambda = 0.3$).

Bibliography

Kusum L Ailawadi, Karen Gedenk, Christian Lutzky, and Scott A Neslin. Decomposition of the sales impact of promotion-induced stockpiling. *Journal of Marketing Research*, pages 450–467, 2007.

D. Bertsimas and R. Weismantel. *Optimization over integers*, volume 13. Dynamic Ideas Belmont, 2005.

Dimitris Bertsimas and Romy Shioda. Algorithm for cardinality-constrained quadratic optimization. *Computational Optimization and Applications*, 43(1):1–22, 2009.

Daniel Bienstock. Computational study of a family of mixed-integer quadratic programming problems. *Mathematical programming*, 74(2):121–140, 1996.

R. C. Blattberg and S. A. Neslin. *Sales promotion: Concepts, Methods, and Strategies*. Prentice Hall, 1990.

Niv Buchbinder, Moran Feldman, Joseph Seffi Naor, and Roy Schwartz. A tight linear time (1/2)-approximation for unconstrained submodular maximization. In *Foundations of Computer Science (FOCS), 2012 IEEE 53rd Annual Symposium on*, pages 649–658. IEEE, 2012.

Gérard P Cachon and Martin A Lariviere. Supply chain coordination with revenue-sharing contracts: strengths and limitations. *Management science*, 51(1):30–44, 2005.

Katia Campo, Els Gijsbrechts, and Patricia Nisol. Towards understanding consumer response to stock-outs. *Journal of Retailing*, 76(2):219–242, 2000.

Felipe Caro and Jérémie Gallien. Clearance pricing optimization for a fast-fashion retailer. *Operations Research*, 60(6):1404–1422, 2012.

Xin Chen, Peng Hu, Stephen Shum, and Yuhan Zhang. Stochastic inventory model with reference price effects. 2013. Submitted.

L. G. Cooper, P. Baron, W. Levy, M. Swisher, and P. Gogos. Promocast: A new forecasting method for promotion planning. *Marketing Science*, 18(3):301–316, 1999.

Daniel Corsten and Thomas Gruen. Stock-outs cause walkouts. *Harvard Business Review*, 82(5):26–28, 2004.

Uriel Feige, Vahab S Mirrokni, and Jan Vondrak. Maximizing non-monotone submodular functions. *SIAM Journal on Computing*, 40(4):1133–1153, 2011.

G. Fibich, A. Gavious, and O. Lowengart. Explicit solutions of optimization models and differential games with nonsmooth (asymmetric) reference-price effects. *Operations Research*, 51(5):721–734, 2003.

Roger Fletcher and Sven Leyffer. Solving mixed integer nonlinear programs by outer approximation. *Mathematical programming*, 66(1-3):327–349, 1994.

Eijte W Foekens, Peter SH Leeflang, and Dick R Wittink. Varying parameter models to accommodate dynamic promotion effects. *Journal of Econometrics*, 89(1):249–268, 1998.

Ignacio E Grossmann. Review of nonlinear mixed-integer and disjunctive programming techniques. *Optimization and Engineering*, 3(3):227–252, 2002.

Sunil Gupta. Impact of sales promotions on when, what, and how much to buy. *Journal of Marketing Research*, pages 342–355, 1988.

H. J. Van Heerde, P. S. H. Leeflang, and D. R. Wittink. The estimation of pre- and post- promotion dips with store-level scanner data. *Journal of Marketing Research*, 37:383–395, 2000.

Raymond Hemmecke, Matthias Köppe, Jon Lee, and Robert Weismantel. Nonlinear integer programming. In *50 Years of Integer Programming 1958-2008*, pages 561–618. Springer, 2010.

P. K. Kopalle, A. G. Rao, and J. L. Assuncao. Asymmetric reference price effects and dynamic pricing policies. *Marketing Science*, 15(1):60–85, 1996.

S. Macé and S. A. Neslin. The determinants of pre- and postpromotion dips in sales of frequently purchased goods. *Journal of Marketing Research*, 41:339–350, 2004.

S Thomas McCormick. Submodular function minimization. *Handbooks in operations research and management science*, 12:321–391, 2005.

Carl F Mela, Kamel Jedidi, and Douglas Bowman. The long-term impact of promotions on consumer stockpiling behavior. *Journal of Marketing Research*, pages 250–262, 1998.

G. L. Nemhauser and L. A. Wolsey. *Integer and combinatorial optimization*, volume 18. Wiley New York, 1988.

George L Nemhauser, Laurence A Wolsey, and Marshall L Fisher. An analysis of approximations for maximizing submodular set functions. *Mathematical Programming*, 14(1):265–294, 1978.

I. Popescu and Y. Wu. Dynamic pricing strategies with reference effects. *Operations Research*, 55(1):98–112, 2007.

E. L. Porteus. Stochastic inventory theory. *Handbooks in OR & MS*, 2:605–652, 1990.

Alexander Schrijver. A combinatorial algorithm minimizing submodular functions in strongly polynomial time. *Journal of Combinatorial Theory, Series B*, 80(2):346–355, 2000.

Hanif D Sherali and Warren P Adams. *A reformulation-linearization technique for solving discrete and continuous nonconvex problems*, volume 31. Springer, 1998.

Shivaram Subramanian and Hanif D Sherali. A fractional programming approach for retail category price optimization. *Journal of Global Optimization*, 48(2):263–277, 2010.

K. T. Talluri and G. J. van Ryzin. *The Theory and Practice of Revenue Management*. Springer, 2005.

K Van Donselaar, Tom Van Woensel, RACM Broekmeulen, and J Fransoo. Inventory control of perishables in supermarkets. *International Journal of Production Economics*, 104(2):462–472, 2006.

J Miguel Villas-Boas. Models of competitive price promotions: Some empirical evidence from the coffee and saltine crackers markets. *Journal of Economics & Management Strategy*, 4(1):85–107, 1995.

Chapter 5

Conclusions

In this thesis, we applied operations research techniques to address three interesting and important practical operational problems. Although there are clear differences that distinguish the three problem settings, I wish to identify the following common themes that they share.

First, one of the key challenges of being a researcher in the field of operations research is that of creating a suitable mathematical model to analyze a complex real-world system. In each of the three settings, we formulated an operational decision as a mathematical optimization problem. In addition, for the case of the Zambia problem, it was necessary for us to build a computer simulation model of the Zambia medical supply chain because the supply chain was too complicated to model using only a mathematical model.

Second, it is often computationally intractable to compute optimal solutions to optimization models that arise from practical operational problems (e.g. computing the optimal promotional price schedule for a grocery retailer). However, with pespiration and inspiration, one can often derive reasonable bounds or heuristics for the problem, sometimes with provable near-optimality guarantees.

Finally, over the course of working on these three problems, I relished the experience of working with real data, as I did not have such an opportunity prior to my doctoral studies at MIT. I believe that mathematical models that are connected to real-world data are more likely to make a positive impact on the world. For instance, we were able to validate our computer simulation model of the Zambia medical supply chain using empirical data collected from health facilities in Zambia. This increases our confidence in the predictions of the computer simulation model.

Although my doctoral thesis is now complete, my hope is to continue to apply operations research techniques to important and challenging problems faced by our world. My desire is also to see our research bear fruit not merely as academic papers, but in improving peoples' health and happiness through improving complex real world systems.