

The Impact of Linear Optimization on Promotion Planning

Maxime C. Cohen

Operations Research Center, MIT, Cambridge, MA 02139, maxcohen@mit.edu

Ngai-Hang Zachary Leung

Operations Research Center, MIT, Cambridge, MA 02139, zacleung@mit.edu

Kiran Panchamgam

Oracle RGBU, Burlington, MA 01803, kiran.panchamgam@oracle.com

Georgia Perakis

Sloan School of Management, MIT, Cambridge, MA 02139, georgiap@mit.edu

Anthony Smith

Oracle RGBU, Burlington, MA 01803, anthony.x.smith@oracle.com

In many important settings, promotions are a key instrument for driving sales and profits. Examples include promotions in supermarkets among others. The Promotion Optimization Problem (POP) is a challenging problem as the retailer needs to decide which products to promote, what is the depth of price discounts and finally, when to schedule the promotions. We introduce and study an optimization formulation that captures several important business requirements as constraints. 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. 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 a linear program (LP) by showing the integrality of the integer program (IP) formulation. We develop analytical results on the accuracy of the approximation relative to the optimal (but intractable) POP solution by showing guarantees relative to the optimal profits. In addition, we show computationally that the formulation solves fast in practice 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 by testing various strategies and business constraints. Finally, we calibrate our models using actual data and determine that they can improve profits by 3% just by optimizing the promotion schedule and up to 5% by slightly modifying some business requirements.

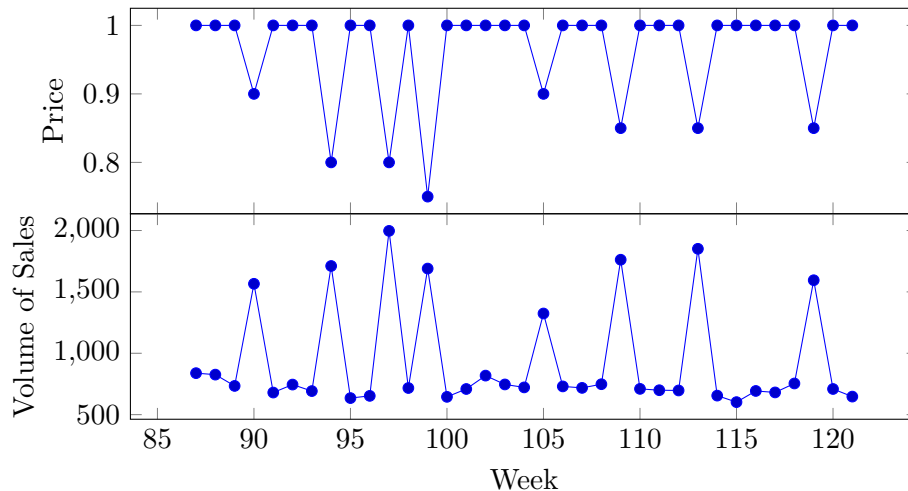
Key words: Promotion Optimization, Dynamic Pricing, Integer Programming, Retail Operations

1. Introduction

Sales promotions have become 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 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 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 1 Prices and sales for a particular brand of coffee during a period of 35 weeks.



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 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 significant.

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.

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 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 3, we describe the model and assumptions we impose as well as the business rules required for our problem. In Section 4, we formulate the Promotion Optimization Problem. In Section 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 6, we consider multiplicative and additive demand models and show bounds on the LP approximation relative to the optimal POP solution. Section 7 presents computational results using real data. Finally, we present our conclusions in Section 8. Several of the proofs of the different propositions and theorems are relegated to the Appendix.

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 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).
- $\mathcal{Q} = \{q^0 > q^1 > \dots > q^k > \dots > q^K\}$ - 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 3.1), one can rewrite the price p_t at time t as follows:

$$p_t = \sum_{k=0}^K q^k \gamma_t^k, \quad (1)$$

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 ; $\forall t = 1, \dots, T$ and $\forall k = 0, \dots, K$, 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 :

$$\sum_{k=0}^K \gamma_t^k = 1; \quad \forall t. \quad (2)$$

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 $p_t, p_{t-1}, \dots, p_{t-M}$ 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 :

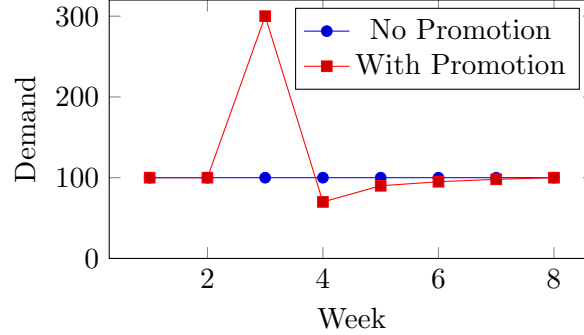
$$d_t(\mathbf{p}_t) = h_t(p_t, p_{t-1}, \dots, p_{t-M}). \quad (3)$$

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

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 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 $p_{t-1}, p_{t-2}, \dots, p_{t-M}$. 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 7).

Figure 2 Example of the promotion fatigue effect. 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 (1), where the price ladder is given by: $q^0 > q^1 > \dots > q^K$. 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:

$$\sum_{t=1}^T \sum_{k=1}^K \gamma_t^k \leq L. \quad (4)$$

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:

$$\sum_{\tau=t}^{t+S} \sum_{k=1}^K \gamma_{\tau}^k \leq 1 \quad \forall t. \quad (5)$$

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 7 and Figure 8). 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 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. Problem Formulation

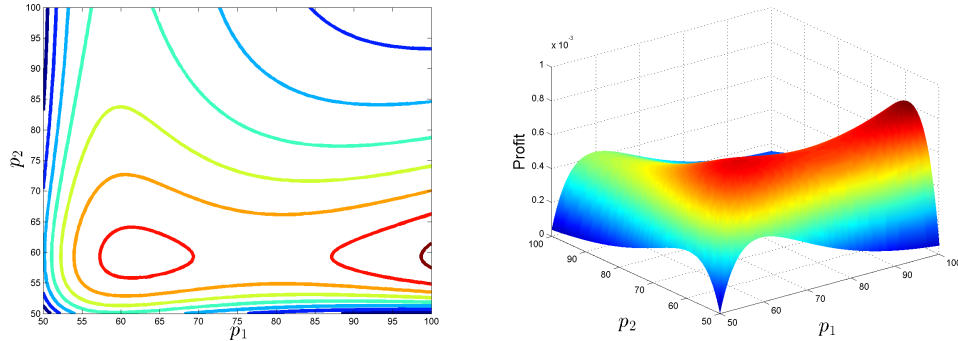
In what follows, we formulate the single-item Promotion Optimization Problem (POP) incorporating the business rules we discussed above:

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

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 $p_{t-1}, \dots, p_{t-M}$ as well as on demand seasonality and trend (see equation (3)). Specific examples are presented in Section 6.

The POP is a nonlinear IP (see Figure 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 nonlinear (neither concave nor convex) due to the cross time dependence between prices (see Figure 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 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.

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 (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 $\mathbf{p} = (p_1, \dots, p_T)$, 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 $\mathbf{p}^0 = (q^0, \dots, q^0)$, for which the regular price is set at all the time periods. Let us define the coefficients b_t^k as:

$$b_t^k = POP(\mathbf{p}_t^k) - POP(\mathbf{p}^0). \quad (6)$$

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:

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

Remark. One can condense the above IP formulation in a more compact way. In particular, since at most one of the decision variables $\{\gamma_t^k : k = 1, \dots, K\}$ is equal to one, one can define $\tilde{b}_t = \max_{k=1, \dots, K} b_t^k$; $\forall t = 1, \dots, T$ 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 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 $t + 2, t + 3, \dots$. 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:

$$\begin{aligned} POP(\mathbf{p}\{p_t = q^i, p_u = q^j\}) - POP(\mathbf{p}^0) = \\ POP(\mathbf{p}\{p_t = q^i\}) - POP(\mathbf{p}^0) + POP(\mathbf{p}\{p_u = q^j\}) - POP(\mathbf{p}^0). \end{aligned} \quad (7)$$

(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. $\square$

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:

$$\left\{ \gamma_t^k : \sum_{t=1}^T \sum_{k=1}^K \gamma_t^k \leq L \quad \forall t; \sum_{\tau=t}^{t+S} \sum_{k=1}^K \gamma_\tau^k \leq 1; \sum_{k=0}^K \gamma_t^k = 1 \quad \forall t \right\}. \quad (8)$$

THEOREM 1. *Every basic feasible solution of (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:

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

Similarly, we denote by $\mathbf{b}$ the vectorization of the objective coefficients b_t^k defined in (6). By relaxing the integrality constraints, the IP problem can be written in the following standard LP form:

$$\begin{aligned} \max_{\gamma} \quad & \mathbf{b}^T \gamma \\ \text{s.t.} \quad & A\gamma \leq \mathbf{u} \\ & 0 \leq \gamma \leq 1 \end{aligned} \quad (9)$$

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:

$$A = \begin{bmatrix} \mathbf{1}_K^T & & & & & & & & & & \\ & \mathbf{1}_K^T & & & & & & & & & \\ & & \ddots & & & & & & & & \\ & & & \ddots & & & & & & & \\ & & & & \ddots & & & & & & \\ & & & & & \ddots & & & & & \\ & & & & & & \ddots & & & & \\ & & & & & & & \mathbf{1}_K^T & & & \\ \hline 0 & \mathbf{1}_K^T & 0 & \mathbf{1}_K^T & \dots & 0 & \mathbf{1}_K^T & 0 & \mathbf{1}_K^T & & \\ & \underbrace{\hspace{10em}}_S & & & & & & & & & \\ & & 0 & \mathbf{1}_K^T & \dots & 0 & \mathbf{1}_K^T & 0 & \mathbf{1}_K^T & 0 & \mathbf{1}_K^T \\ & & \underbrace{\hspace{10em}}_S & & & & & & & & \\ & & & \ddots & & & & & & & \\ & & & & 0 & \mathbf{1}_K^T & 0 & \mathbf{1}_K^T & \dots & 0 & \mathbf{1}_K^T \\ & & & & \underbrace{\hspace{10em}}_S & & & & & & \\ \hline 0 & \mathbf{1}_K^T & 0 & \mathbf{1}_K^T & \dots & \dots & \dots & \dots & 0 & \mathbf{1}_K^T \end{bmatrix}; \quad \mathbf{u} = \begin{bmatrix} e_T \\ e_{T-S+1} \\ L \end{bmatrix}.$$

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 $t = 1, 2, \dots, T$. We note that in (9), the equality is transformed to an inequality.

This can be done because $b_t^0 = 0$ for all $t = 1, 2, \dots, T$. 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 (5). Finally, the last row of A corresponds to the constraint on the limitation on the number of promotions allowed from (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:

$$\tilde{A} = \begin{bmatrix} \underbrace{\mathbf{1}_K^T \mathbf{1}_K^T \dots \mathbf{1}_K^T \mathbf{1}_K^T}_S & & & \\ & \underbrace{\mathbf{1}_K^T \dots \mathbf{1}_K^T \mathbf{1}_K^T \mathbf{1}_K^T}_S & & \\ & & \ddots & \\ & & & \underbrace{\mathbf{1}_K^T \mathbf{1}_K^T \dots \mathbf{1}_K^T}_S \\ \mathbf{1}_K^T \mathbf{1}_K^T \dots & & & \mathbf{1}_K^T \end{bmatrix}.$$

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 (8) is integral. $\square$

Using Theorem 1, one can solve (IP) efficiently by solving its LP relaxation, given by:

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

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.

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 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 (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 (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 (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 1). The estimated demand model for this coffee brand follows the following relation:

$$\log d_t = \beta^0 + \beta^1 t + \beta^2 WEEK_t - 3.277 \log p_t + 0.518 \log p_{t-1} + 0.465 \log p_{t-2}. \quad (10)$$

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

Table 1 Forecast accuracy of two regression models for a brand of coffee.

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

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 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 (10).

Notation We introduce the following notation that will be used in the sequel. Let $A = \{(t_1, k_1), \dots, (t_N, k_N)\}$ with $N \leq L$ be a set of promotions with $1 \leq t_1 < t_2 < \dots < t_N \leq T$. In other words, at each time period t_n ; $\forall n = 1, \dots, N$ 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:

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

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

Example. Suppose that the price ladder is given by $\mathcal{Q} = \{q^0 = 5 > q^1 = 4 > q^2 = 3\}$, 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:

$$LP(\gamma) = POP(\mathbf{p}^0) + \sum_{t=1}^T \sum_{k=1}^K b_t^k \gamma_t^k, \quad (11)$$

where b_t^k is defined in (6). Finally, we denote by $\tilde{L}$ the effective maximal number of promotions given by:

$$\tilde{L} = \min\{L, \tilde{N}\}, \quad \text{where } \tilde{N} = \left\lfloor \frac{T-1}{S+1} \right\rfloor + 1. \quad (12)$$

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 $\tilde{L} \geq 1$.

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:

$$d_t = f_t(p_t) \cdot g_1(p_{t-1}) \cdot g_2(p_{t-2}) \cdots g_M(p_{t-M}). \quad (13)$$

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 7 from the actual data, it is reasonable to assume the following for the functions g_k .

- ASSUMPTION 1. 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. The demand in (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:

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

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 (13).

6.1.1. Bounds on Quality of Approximation

THEOREM 2. *Let γ^{POP} be an optimal solution to (POP) and let γ^{LP} be an optimal solution to (LP). Then:*

$$1 \leq \frac{POP(\gamma^{POP})}{POP(\gamma^{LP})} \leq \frac{1}{\underline{R}}, \quad (14)$$

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

$$\underline{R} = \prod_{i=1}^{\tilde{L}-1} g_{i(S+1)}(q^K), \quad (15)$$

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:

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

Inequality (i) follows from Proposition 2 below. Inequality (ii) follows from the optimality of γ^{POP} and inequality (iii) follows from part 2 of Lemma 1 below. Finally, inequality (iv) follows from the optimality of γ^{LP} . Therefore, we obtain:

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

Theorem 2 relies on the following two results.

LEMMA 1 (Submodular effect of the last promotion on profits).

1. Let $A = \{(t_1, k_1), \dots, (t_n, k_n)\}$ be a set of promotions with $t_1 < t_2 < \dots < t_n$ ($n \leq L$) 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:

$$POP(\gamma_{A \cup \{(t', k')\}}) \geq POP(\gamma_A), \quad (16)$$

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

$$POP(\gamma_{A \cup \{(t', k')\}}) - POP(\gamma_A) \leq POP(\gamma_{B \cup \{(t', k')\}}) - POP(\gamma_B). \quad (17)$$

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

Note that if (16) is not satisfied, the sub-additivity property of Lemma 1 does not necessarily hold for any feasible solution. However, the required condition in (16) is always automatically satisfied for the optimal POP solution. The proof of Lemma 1 can be found in Appendix A. Lemma 1 states that for a multiplicative demand model as in (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 2. For any feasible vector γ , we have: $POP(\gamma) \geq \underline{R} \cdot LP(\gamma)$.

The proof of Proposition 2 can be found in Appendix 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 2, 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 (13). In addition, as we illustrate in Section 6.1.2, these bounds perform well in practice for a wide range of parameters.

We next show that the bounds of Theorem 2 are tight.

PROPOSITION 3 (Tightness of the bounds for multiplicative demand).

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

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

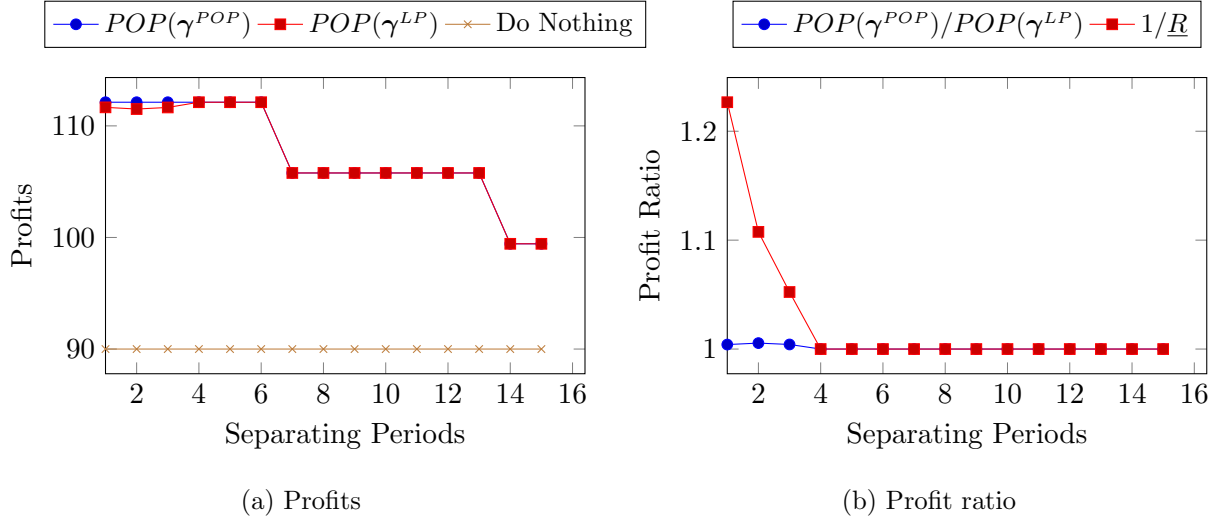
The proof of Proposition 3 can be found in Appendix C.

6.1.2. 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 2 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 2.

In Figures 4, 5 and 6, the demand model we use is given by: $\log d_t(\mathbf{p}) = \log(10) - 4 \log p_t + 0.5 \log p_{t-1} + 0.3 \log p_{t-2} + 0.2 \log p_{t-3} + 0.1 \log p_{t-4}$.

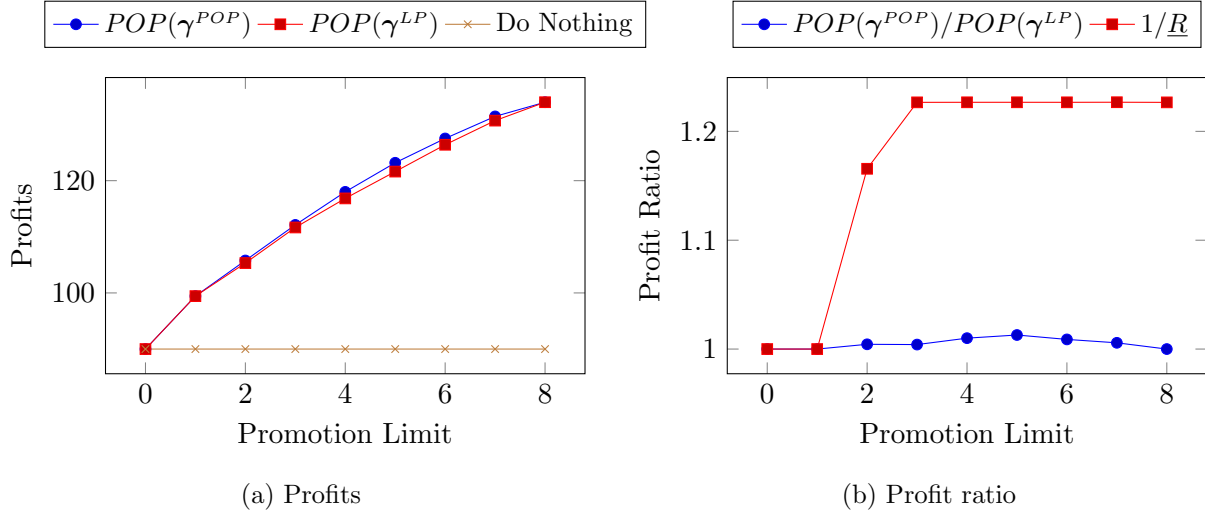
Figure 4 Effect of varying the separating parameter S .

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

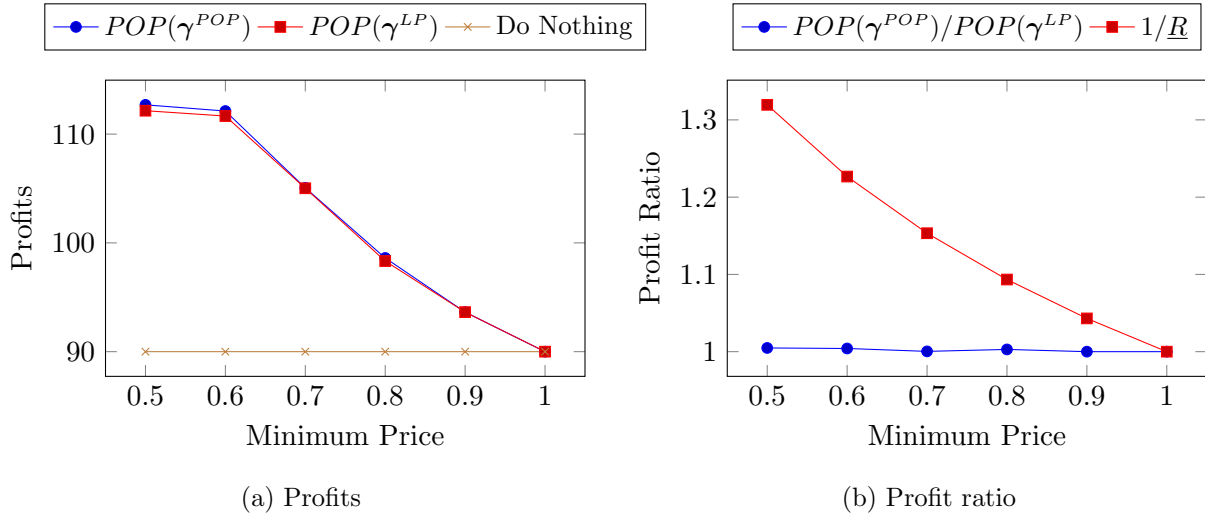
Dependence on separating periods: In Figure 4, we vary the number of separating periods S from 1 to 16 (remember that the horizon is $T = 35$ weeks). We make the following observations: a) As one would expect from Proposition 1, the LP approximation coincides with the optimal POP solution when $S \geq M = 4$, 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 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 5, we vary the number of promotions allowed L between 0 and 8. We make the following observations: a) As one would expect from Proposition 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 (35) of Theorem 2, $1/\underline{R}$ 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 6, 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 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 (35), $1/\underline{R}$ decreases with q^K .

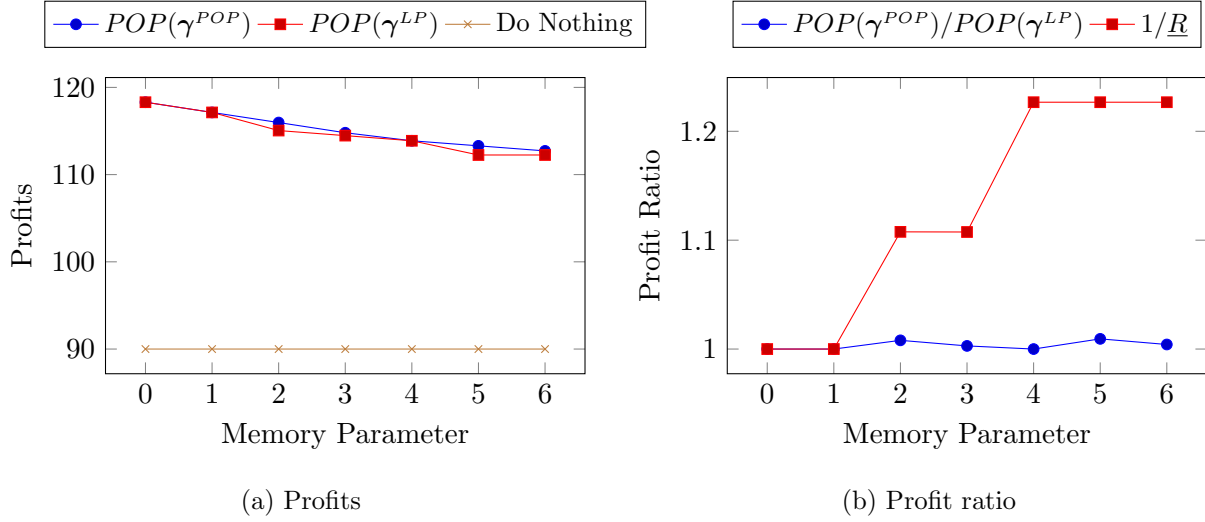
Figure 5 Effect of varying the number of promotions allowed L .

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

Figure 6 Effect of varying the minimum price q^K 

Note. Example parameters: $L = 3$, $S = 1$.

Dependence on the length of the memory: In Figure 7, 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 $g_1, g_2, \dots, g_M$ 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 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 (35), $1/\underline{R}$ increases with M .

Figure 7 Effect of varying the memory parameter M 

Note. Example parameters: $\log d_t(\mathbf{p}) = \log(10) - 4\log p_t + 0.2\log p_{t-1} + 0.2\log p_{t-2} + \dots + 0.2\log p_{t-M}$; $L = 3, S = 1$.

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 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:

$$d_t = f_t(p_t) + g_1(p_{t-1}) + g_2(p_{t-2}) + \dots + g_M(p_{t-M}). \quad (18)$$

As we verify in Section 7 from the actual data, it is reasonable to assume the following structure for the functions g_k .

- ASSUMPTION 2. 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 1 for the multiplicative model. We assume that for $k > M$, $g_k(p) = 0 \forall p$.

Remark. Equation (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:

$$d_t = a - \delta p_t - \phi(p_t - r_t). \quad (19)$$

Equation (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 (19) can be written as follows in terms of the current and past prices:

$$d_t = a - (\delta + \phi)p_t + \sum_{k=1}^{M=T} (1 - \theta)\phi \theta^{k-1} p_{t-k}. \quad (20)$$

One can see that equation (20) falls under the model we proposed in (18), when the functions g_k are chosen appropriately and the memory parameter M goes to infinity. In addition, the additive model from (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 (18).

6.2.1. Bounds on Quality of Approximation

THEOREM 3. *Let γ^{POP} be an optimal solution to (POP) and let γ^{LP} be an optimal solution to the LP approximation. Then:*

$$1 \leq \frac{POP(\gamma^{POP})}{POP(\gamma^{LP})} \leq 1 + \frac{\bar{R}}{POP(\gamma^{LP})}. \quad (21)$$

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

$$\bar{R} = \sum_{i=1}^{\bar{L}} \sum_{j=i+1}^{\bar{L}} (q^K - q^0) g_{(j-i)(S+1)}(q^K). \quad (22)$$

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:

$$LP(\gamma^{LP}) \stackrel{(i)}{\leq} POP(\gamma^{LP}) \stackrel{(ii)}{\leq} POP(\gamma^{POP}) \stackrel{(iii)}{\leq} LP(\gamma^{POP}) + \bar{R} \stackrel{(iv)}{\leq} LP(\gamma^{LP}) + \bar{R}. \quad (23)$$

Inequalities (i) and (iii) follow from Proposition 4 below. Inequality (ii) follows from the optimality of γ^{POP} and inequality (iv) follows from the optimality of γ^{LP} . Therefore, we obtain:

$$1 = \frac{POP(\gamma^{LP})}{POP(\gamma^{LP})} \leq \frac{POP(\gamma^{POP})}{POP(\gamma^{LP})} \leq \frac{LP(\gamma^{LP}) + \bar{R}}{POP(\gamma^{LP})} \leq \frac{POP(\gamma^{LP}) + \bar{R}}{POP(\gamma^{LP})} = 1 + \frac{\bar{R}}{POP(\gamma^{LP})}. \quad \square$$

The proof of Theorem 3 relies on the following result.

PROPOSITION 4. *For a given promotion profile γ , with the promotion set: $\{(t_1, k_1), \dots, (t_n, k_n)\}$, the POP profits can be written as follows:*

$$POP(\gamma_{\{(t_1, k_1), \dots, (t_n, k_n)\}}) = LP(\gamma_{\{(t_1, k_1), \dots, (t_n, k_n)\}}) + ER(\gamma_{\{(t_1, k_1), \dots, (t_n, k_n)\}}). \quad (24)$$

Here, $ER(\gamma_{\{(t_1, k_1), \dots, (t_n, k_n)\}})$ represents the error term between the POP and the LP objectives and is given by:

$$ER(\gamma_{\{(t_1, k_1), \dots, (t_n, k_n)\}}) = \sum_{i=1}^n \sum_{j=i+1}^n (q^{k_j} - q^0) g_{t_j - t_i}(q^{k_i}). \quad (25)$$

Consequently, for any feasible promotion profile γ , the POP profits satisfies:

$$LP(\gamma) \leq POP(\gamma) \leq LP(\gamma) + \bar{R}.$$

The proof of Proposition 4 can be found in Appendix D. Proposition 4 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 $A = \{(t_1, k_1), \dots, (t_N, k_N)\}$ be a set of promotions with $1 \leq t_1 < t_2 < \dots < t_n$ ($n \leq L$) 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:

$$POP(\gamma_{A \cup \{(t', k')\}}) - POP(\gamma_A) \geq POP(\gamma_{B \cup \{(t', k')\}}) - POP(\gamma_B). \quad (26)$$

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 (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 (24) in Proposition 4, we obtain:

$$\begin{aligned} POP(\gamma_A) &= LP(\gamma_A) + \sum_{(t,k),(u,\ell) \in A: t < u} (q^\ell - q^0) g_{u-t}(q^\ell) \\ POP(\gamma_{A \cup \{(t',k')\}}) &= LP(\gamma_{A \cup \{(t',k')\}}) + \sum_{(t,k),(u,\ell) \in A \cup \{(t',k')\}: t < u} (q^\ell - q^0) g_{u-t}(q^\ell), \end{aligned}$$

and similarly for the set B . By using the definition of the LP objective function:

$$LP(\gamma_{\{(t_1,k_1), \dots, (t_n,k_n)\}}) = POP(\gamma^0) + \sum_{i=1}^n (POP(\gamma_{\{t_i,k_i\}}) - POP(\gamma^0)),$$

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:

$$\begin{aligned} POP(\gamma_{A \cup \{(t',k')\}}) - POP(\gamma_A) &= POP(\gamma') - POP(\gamma^0) + \sum_{(t,k) \in A} \phi((t,k), (t',k')), \\ POP(\gamma_{B \cup \{(t',k')\}}) - POP(\gamma_B) &= POP(\gamma') - POP(\gamma^0) + \sum_{(t,k) \in B} \phi((t,k), (t',k')). \end{aligned}$$

Therefore, we obtain:

$$\begin{aligned} \left(POP(\gamma_{A \cup \{(t',k')\}}) - POP(\gamma_A) \right) - \left(POP(\gamma_{B \cup \{(t',k')\}}) - POP(\gamma_B) \right) \\ = \sum_{(t,k) \in A \setminus B} \phi((t,k), (t',k')) \geq 0. \quad \square \end{aligned}$$

Corollary 1 states that for an additive demand model as in (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 $p_1, p_2, \dots, p_T$, 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 3 are tight.

PROPOSITION 5 (Tightness of the bounds for additive model).

1. *The lower bound in Theorem 3 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 3 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:*

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

The proof can be found in Appendix E.

6.2.2. Illustrating the bounds. For brevity, the plots where we illustrate the bounds for the additive demand model are presented in Appendix F. We refer the reader to Section 6.1.2 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.

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:

$$d_t = \lambda \cdot d_1(p_t, p_{t-1}, \dots, p_{t-M}) + (1 - \lambda) \cdot d_2(p_t, p_{t-1}, \dots, p_{t-M}), \quad (27)$$

where $d_1(p_t, p_{t-1}, \dots, p_{t-M})$ is a multiplicative model as in (13) and $d_2(p_t, p_{t-1}, \dots, p_{t-M})$ is an additive model as in (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 (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 (27). Note that if $\lambda = 0$, (27) reduces to the additive class of demand functions we discussed in Section 6.2; whereas if $\lambda = 1$, (27) reduces to the multiplicative class of demand functions we discussed in Section 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 (27), one can still naturally use the LP approximation method described in Section 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 (27). We note that for the unified demand model in (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:

$$\Pi = \max \{POP_1(\gamma^{LP_1}), POP_2(\gamma^{LP_2}), POP(\gamma^{LP_{unif}})\}, \quad (28)$$

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 (28) are feasible to the POP for the unified demand model, we obtain:

$$\Pi \leq POP(\gamma^{POP_{unif}}), \quad (29)$$

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 (27) are presented in the following Theorem.

THEOREM 4. *Let $\gamma^{POP_{unif}}$ be an optimal solution to (POP), and let Π be defined as in (28). Then:*

$$1 \leq \frac{POP(\gamma^{POP_{unif}})}{\Pi} \leq UB2 = \frac{\lambda}{\underline{R}_1} + (1 - \lambda) \cdot \left[1 + \frac{\overline{R}_2}{POP_2(\gamma^{LP_2})}\right], \quad (30)$$

where $\underline{R}_1$ and $\overline{R}_2$ are given by (35) and (22) respectively.

Proof. The first inequality follows directly from equation (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:

$$\lambda \cdot POP_1(\gamma^{POP_1}) + (1 - \lambda) \cdot POP_2(\gamma^{POP_2}) \leq \frac{\lambda}{\underline{R}_1} \cdot POP_1(\gamma^{LP_1}) + (1 - \lambda) \cdot \left[1 + \frac{\overline{R}_2}{POP_2(\gamma^{LP_2})} \right] \cdot POP_2(\gamma^{LP_2}).$$

The proof can be concluded by using the definition of Π . $\square$

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:*

$$POP(\gamma^{LP_{unif}}) \leq POP(\gamma^{POP_{unif}}) \leq UB1 = \lambda \cdot LP_1(\gamma^{LP_1}) + (1 - \lambda) \cdot \left[LP_2(\gamma^{LP_2}) + \overline{R}_2 \right], \quad (31)$$

where $\overline{R}_2$ is given by (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:

$$\lambda \cdot POP_1(\gamma^{POP_1}) + (1 - \lambda) \cdot POP_2(\gamma^{POP_2}) \leq LP_1(\gamma^{LP_1}) + (1 - \lambda) \cdot \left[LP_2(\gamma^{LP_2}) + \overline{R}_2 \right]. \quad \square$$

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 (27) one can compute guarantees on the performance given in equation (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 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 6.1 and 6.2 respectively.

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 (13) and often used in practice.

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 (10)):

$$\log d_{it} = \beta^0 BRAND_i + \beta^1 t + \beta^2 WEEK_t + \sum_{m=0}^M \beta_{im}^3 \log p_{i,t-m} + \epsilon_t, \quad (32)$$

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 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 ; $\forall t = 1, \dots, T$ 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 (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:

$$\begin{aligned} f_t(p_t) &= \exp(\beta^0 + \beta^1 t + \beta^2 WEEK_t + \beta_0^3 \log p_t), \\ g_m(p_{t-m}) &= (p_{t-m})^{\beta_m^3}, \quad m = 1, \dots, M, \end{aligned}$$

and therefore, equation (32) is in fact a special case of the multiplicative demand model in (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 $m = 1, \dots, M - 1$.

We note that the conditions above are a special case of Assumption 1 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:

$$MAPE = \frac{1}{T} \sum_{t=1}^T \frac{|s_t - \hat{s}_t|}{s_t}.$$

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:

$$R^2 = 1 - \frac{SS_{res}}{SS_{tot}},$$

Table 2 A subset of the estimation results for two coffee brands.

	Variable	Coefficient	Std Error	p-value
Brand1	$\log p_t$	-3.277	0.231	2e-16***
	$\log p_{t-1}$	0.518	0.229	0.024*
	$\log p_{t-2}$	0.465	0.231	0.045*
Brand2	$\log p_t$	-4.434	0.427	2e-16***
	$\log p_{t-1}$	1.078	0.423	0.011*
	$\log p_{t-2}$	0.067	0.413	0.870
Brand1	MAPE		0.116	
	OOS R^2		0.900	
	Revenue bias		1.059	
Brand2	MAPE		0.097	
	OOS R^2		0.903	
	Revenue bias		1.017	

In-sample adjusted R^2 Significance codes: * indicates significance < 0.05 ,
 *** indicates significance < 0.001 .

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:

$$R_{adj}^2 = 1 - (1 - R^2) \cdot \frac{n - 1}{n - p - 1},$$

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:

$$\text{revenue bias} = \frac{\sum_{t=1}^T p_t \hat{s}_t}{\sum_{t=1}^T p_t s_t}.$$

7.2. Estimation Results and Discussion

7.2.1. Coffee Category The coffee category is an appropriate candidate to test our model as it is common in promotion applications (see e.g., ? and ?). We use a linear regression to estimate the parameters of the demand model in equation (32) for five different brands of coffee. For conciseness, we only present a subset of the estimation results for two coffee brands in Table 2. We compare the actual and predicted sales in Figure 8. Remember that our data consists of 117 weeks which we split into 82 weeks on training and 35 weeks of testing.

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 $M = 2$.

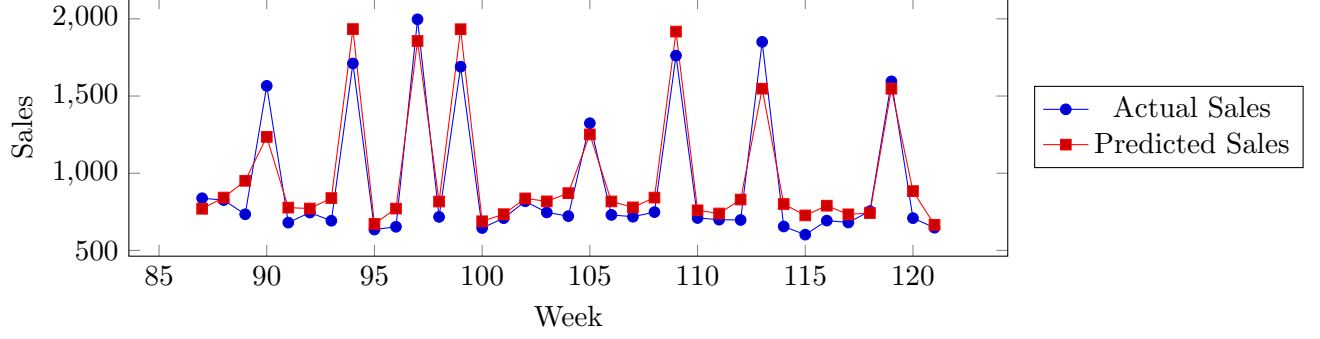


Figure 8 Actual versus forecasted sales over the 35 test weeks for Brand1.

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.

7.2.2. 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 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.

Table 3 Summary of the estimation results for four product categories.

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

The column for Product Memories indicates all values of the memory parameter that we found for the products in that category.

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.

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 $\{0.75, 0.80, 0.85, \dots, 1\}$.

The LP optimization results are shown in Figure 9. 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.

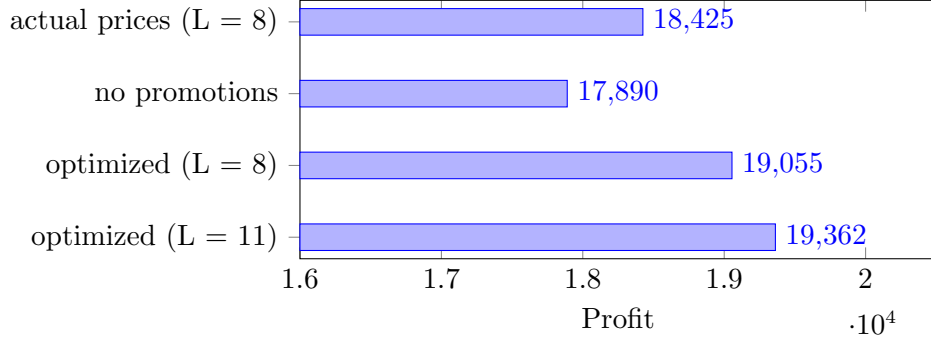


Figure 9 Profits for different scenarios using a log-log demand model.

We next compute the bound from Theorem 2 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 $\underline{R} = \prod_{i=1}^{\bar{L}-1} g_{i(S+1)}(q^K)$ and therefore, depends on the parameters of the problem. We compute $\underline{R}$ for both coffee brands from Table 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 10. 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.

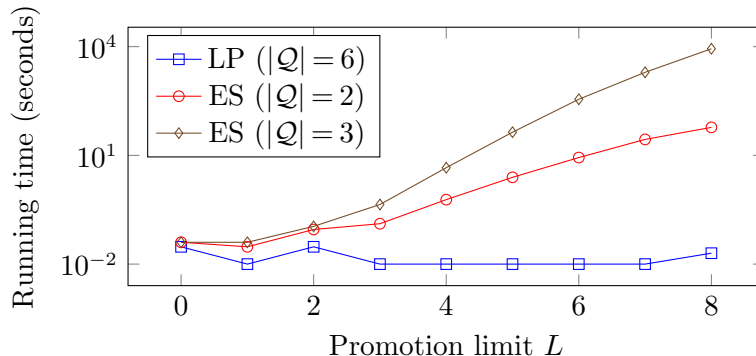


Figure 10 Running times of the LP formulation and the exhaustive search method for the POP. The number of the price ladder elements is $|\mathcal{Q}|$. The LP was solved using the Java interface to Gurobi 5.5.0.

The above results show that the exhaustive search method is clearly not a viable option in practice. Note that the LP formulation solves very fast. An important feature of our method relies on the fact that in practice, one can implement it on a platform such as Excel. For a category manager in charge of around 300 SKUs, solving the POP for each item independently would require only about 15 seconds. An additional advantage of short running times is that it allows category managers to perform a sensitivity analysis with respect to the business requirements and to the model parameters. For example, if the optimization is embedded into a decision-support tool, category managers could perform interactive “what-if” analysis. In practice, this would not be possible in the case where the optimization running times exceed a few minutes. In addition, as we have shown in this paper, the LP formulation yields a solution that is accurate relative to the POP optimal prices and one can compute the upper and lower bounds as a guarantee.

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.

References

- Ailawadi, Kusum L, Karen Gedenk, Christian Lutzky, Scott A Neslin. 2007. Decomposition of the sales impact of promotion-induced stockpiling. *Journal of Marketing Research* 450–467.
- Bertsimas, D., R. Weismantel. 2005. *Optimization over integers*, vol. 13. Dynamic Ideas Belmont.
- Bertsimas, Dimitris, Romy Shioda. 2009. Algorithm for cardinality-constrained quadratic optimization. *Computational Optimization and Applications* **43**(1) 1–22.
- Bienstock, Daniel. 1996. Computational study of a family of mixed-integer quadratic programming problems. *Mathematical programming* **74**(2) 121–140.
- Blattberg, R. C., S. A. Neslin. 1990. *Sales promotion: Concepts, Methods, and Strategies*. Prentice Hall.
- Buchbinder, Niv, Moran Feldman, Joseph Seffi Naor, Roy Schwartz. 2012. A tight linear time $(1/2)$ -approximation for unconstrained submodular maximization. *Foundations of Computer Science (FOCS), 2012 IEEE 53rd Annual Symposium on*. IEEE, 649–658.
- Cachon, Gérard P, Martin A Lariviere. 2005. Supply chain coordination with revenue-sharing contracts: strengths and limitations. *Management science* **51**(1) 30–44.
- Campo, Katia, Els Gijsbrechts, Patricia Nisol. 2000. Towards understanding consumer response to stock-outs. *Journal of Retailing* **76**(2) 219–242.
- Caro, Felipe, Jérémie Gallien. 2012. Clearance pricing optimization for a fast-fashion retailer. *Operations Research* **60**(6) 1404–1422.
- Chen, Xin, Peng Hu, Stephen Shum, Yuhang Zhang. 2013. Stochastic inventory model with reference price effects. Submitted.
- Cooper, L. G., P. Baron, W. Levy, M. Swisher, P. Gogos. 1999. Promocast: A new forecasting method for promotion planning. *Marketing Science* **18**(3) 301–316.
- Corsten, Daniel, Thomas Gruen. 2004. Stock-outs cause walkouts. *Harvard Business Review* **82**(5) 26–28.
- Feige, Uriel, Vahab S Mirrokni, Jan Vondrak. 2011. Maximizing non-monotone submodular functions. *SIAM Journal on Computing* **40**(4) 1133–1153.
- Fibich, G., A. Gavious, O. Lowengart. 2003. Explicit solutions of optimization models and differential games with nonsmooth (asymmetric) reference-price effects. *Operations Research* **51**(5) 721–734.

- Fletcher, Roger, Sven Leyffer. 1994. Solving mixed integer nonlinear programs by outer approximation. *Mathematical programming* **66**(1-3) 327–349.
- Foekens, Eijte W, Peter SH Leeftang, Dick R Wittink. 1998. Varying parameter models to accommodate dynamic promotion effects. *Journal of Econometrics* **89**(1) 249–268.
- Grossmann, Ignacio E. 2002. Review of nonlinear mixed-integer and disjunctive programming techniques. *Optimization and Engineering* **3**(3) 227–252.
- Heerde, H. J. Van, P. S. H. Leeftang, D. R. Wittink. 2000. The estimation of pre- and post- promotion dips with store-level scanner data. *Journal of Marketing Research* **37** 383–395.
- Hemmecke, Raymond, Matthias Köppe, Jon Lee, Robert Weismantel. 2010. Nonlinear integer programming. *50 Years of Integer Programming 1958-2008*. Springer, 561–618.
- Kopalle, P. K., A. G. Rao, J. L. Assuncao. 1996. Asymmetric reference price effects and dynamic pricing policies. *Marketing Science* **15**(1) 60–85.
- Macé, S., S. A. Neslin. 2004. The determinants of pre- and postpromotion dips in sales of frequently purchased goods. *Journal of Marketing Research* **41** 339–350.
- McCormick, S Thomas. 2005. Submodular function minimization. *Handbooks in operations research and management science* **12** 321–391.
- Mela, Carl F, Kamel Jedidi, Douglas Bowman. 1998. The long-term impact of promotions on consumer stockpiling behavior. *Journal of Marketing Research* 250–262.
- Nemhauser, G. L., L. A. Wolsey. 1988. *Integer and combinatorial optimization*, vol. 18. Wiley New York.
- Nemhauser, George L, Laurence A Wolsey, Marshall L Fisher. 1978. An analysis of approximations for maximizing submodular set functions. *Mathematical Programming* **14**(1) 265–294.
- Popescu, I., Y. Wu. 2007. Dynamic pricing strategies with reference effects. *Operations Research* **55**(1) 98–112.
- Porteus, E. L. 1990. Stochastic inventory theory. *Handbooks in OR & MS* **2** 605–652.
- Schrijver, Alexander. 2000. A combinatorial algorithm minimizing submodular functions in strongly polynomial time. *Journal of Combinatorial Theory, Series B* **80**(2) 346–355.
- Sherali, Hanif D, Warren P Adams. 1998. *A reformulation-linearization technique for solving discrete and continuous nonconvex problems*, vol. 31. Springer.
- Subramanian, Shivaram, Hanif D Sherali. 2010. A fractional programming approach for retail category price optimization. *Journal of Global Optimization* **48**(2) 263–277.
- Talluri, K. T., G. J. van Ryzin. 2005. *The Theory and Practice of Revenue Management*. Springer.
- Van Donselaar, K, Tom Van Woensel, RACM Broekmeulen, J Fransoo. 2006. Inventory control of perishables in supermarkets. *International Journal of Production Economics* **104**(2) 462–472.

Appendix A: Proof of Lemma 1

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 1, we have $0 \leq \delta_{t'} \leq \delta_{t'+1} \leq \cdots \leq \delta_T \leq 1$. In addition, we have: $a_t = \delta_t b_t$, $a'_t = \delta_t b'_t$. Observe also that condition (16) is equivalent to:

$$\sum_{t=1}^T a'_t - \sum_{t=1}^T a_t \geq 0. \quad (33)$$

Note that $a_t = a'_t$ for all $t < t'$. In the example, we have $a_1 = a'_1, \dots, a_4 = a'_4$ as the prices in periods 1-4 are the same. Therefore, (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 (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: $0.1(b_6 - b'_6) \leq 0.2(b_6 - b'_6) \leq 0.2(b'_5 - b_5)$. $\square$

2. We first introduce the following notation. Let γ^{POP} be an optimal solution to the POP and $\{(t_1, k_1), \dots, (t_n, k_n)\}$ the set of promotions in γ^{POP} . For any subset $B \subset \{1, 2, \dots, n\}$, 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:

$$POP(\gamma^{POP}) = POP(\gamma\{1\}) + \sum_{m=1}^{n-1} [POP(\gamma\{1, \dots, m+1\}) - POP(\gamma\{1, \dots, m\})].$$

Based on Proposition 6 below, we have for each $m = 1, 2, \dots, n-1$: $POP(\gamma\{1, \dots, m+1\}) - POP(\gamma\{1, \dots, m\}) \geq 0$. By applying the submodularity property from Lemma 1 part 1, we obtain: $0 \leq POP(\gamma\{1, \dots, m+1\}) - POP(\gamma\{1, \dots, m\}) \leq POP(\gamma\{m+1\}) - POP(\gamma^0)$. Therefore, we have:

$$\begin{aligned} POP(\gamma^{POP}) &= POP(\gamma\{1\}) + \sum_{m=1}^{n-1} [POP(\gamma\{1, \dots, m+1\}) - POP(\gamma\{1, \dots, m\})] \\ &\leq POP(\gamma^0) + \sum_{m=1}^n [POP(\gamma\{m\}) - POP(\gamma^0)] = LP(\gamma^{POP}). \end{aligned}$$

PROPOSITION 6. *Let $n \geq 2$ be an integer and γ^{POP} an optimal solution to the POP with n promotions. Then, $POP(\gamma\{1, \dots, m+1\}) - POP(\gamma\{1, \dots, m\}) \geq 0$ for $m = 1, 2, \dots, n-1$.*

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 $\{(t_2, k_2), \dots, (t_{n+1}, k_{n+1})\}$ is an optimal solution to POP' with n promotions. Therefore, by using the induction hypothesis, we have:

$$\begin{array}{ccc} POP'(\gamma\{2, \dots, n, n+1\}) & - POP'(\gamma\{2, \dots, n\}) & \geq 0 \\ \vdots & \vdots & \vdots \\ POP'(\gamma\{2, 3\}) & - POP'(\gamma\{2\}) & \geq 0 \end{array}$$

Equivalently, in terms of the POP:

$$\begin{array}{ccc} POP(\gamma\{1, \dots, n, n+1\}) & - POP(\gamma\{1, \dots, n\}) & \geq 0 \\ \vdots & \vdots & \vdots \\ POP(\gamma\{1, 2, 3\}) & - POP(\gamma\{1, 2\}) & \geq 0 \end{array}$$

Therefore, it remains to show that: $POP(\gamma\{1, 2\}) - POP(\gamma\{1\}) \geq 0$. We next prove the following chain of inequalities:

$$\begin{aligned} POP(\gamma\{1, 2\}) - POP(\gamma\{1\}) &\geq POP(\gamma\{1, 2, 3\}) - POP(\gamma\{1, 3\}) \\ &\geq POP(\gamma\{1, 2, 3, 4\}) - POP(\gamma\{1, 3, 4\}) \\ &\vdots \\ &\geq POP(\gamma\{1, \dots, n, n+1\}) - POP(\gamma\{1, 3, 4, \dots, n, n+1\}). \end{aligned} \tag{34}$$

By using the induction hypothesis together with the submodularity property from Lemma 1 part 1, we obtain for each $m = 2, 3, \dots, n-1$:

$$\begin{aligned} POP(\gamma\{1, \dots, m, m+1\}) - POP(\gamma\{1, \dots, m\}) &\leq \\ &POP(\gamma\{1, 3, 4, \dots, m, m+1\}) - POP(\gamma\{1, 3, 4, \dots, m\}). \end{aligned}$$

Finally, from the optimality of $\gamma\{1, \dots, n, n+1\}$ for the POP, we have: $POP(\gamma\{1, \dots, n, n+1\}) - POP(\gamma\{1, \dots, n\}) \geq 0$. By rearranging the terms in the above equations, one can derive the chain of inequalities in (34) and this concludes the proof. $\square$

Appendix B: Proof of Proposition 2

Proof. We denote the set of promotions in the price vector $\mathbf{p}$ by: $\mathbf{p} = \mathbf{p}\{(t_1, q^{t_1}), \dots, (t_N, q^{t_N})\}$, where N is the number of promotions. The price vector $\mathbf{p}^n = \mathbf{p}\{(t_n, q^{t_n})\}$ for each $n = 1, \dots, N$ 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 $\mathbf{p}^0 = (q^0, \dots, q^0)$. We denote the cumulative POP objective in periods $[u, v)$ when using $\mathbf{p}^n$ by:

$$x_{[u,v]}^n = POP(\mathbf{p}\{(t_n, q^{t_n})\})_{[u,v]} = \sum_{t=u}^{v-1} \mathbf{p}_t\{(t_n, q^{t_n})\} d_t(\mathbf{p}_t\{(t_n, q^{t_n})\}).$$

Note that the LP objective can be written as: $LP(\mathbf{p}) = x_{[1,T]}^0 + \sum_{n=1}^N (x_{[1,T]}^n - x_{[1,T]}^0)$.

Since $\mathbf{p}^n$ and $\mathbf{p}^0$ do not promote before t_n , we have $x_{[1,t_n]}^n = x_{[1,t_n]}^0$. 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., $x_{[t_{n+1},T]}^n \leq x_{[t_{n+1},T]}^0$. Therefore, we obtain for each $n = 1, \dots, N$:

$$x_{[1,T]}^n - x_{[1,T]}^0 = x_{[1,t_n]}^n + x_{[t_n,t_{n+1}]}^n + x_{[t_{n+1},T]}^n - x_{[1,t_n]}^0 - x_{[t_n,t_{n+1}]}^0 - x_{[t_{n+1},T]}^0 \leq x_{[t_n,t_{n+1}]}^n - x_{[t_n,t_{n+1}]}^0.$$

Therefore: $LP(\mathbf{p}) \leq UB = x_{[1,T]}^0 + \sum_{n=1}^N (x_{[t_n,t_{n+1}]}^n - x_{[t_n,t_{n+1}]}^0) = x_{[1,t_1]}^0 + \sum_{n=1}^N x_{[t_n,t_{n+1}]}^n$.

Let UB_t denote the value of UB at time t . Specifically, if $t \in [t_n, t_{n+1})$, then $UB_t = x_t^n$. 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: $\underline{R} \cdot LP(\mathbf{p}) \leq \underline{R} \cdot UB \leq POP(\mathbf{p})$.

□

Appendix C: Proofs of Tightness for Multiplicative Demand

1. Lower bound

Proof. In the case when $S \geq M$, we know from Proposition 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:

$$\mathbf{p}^* = \left(\underbrace{q^K, q^0, \dots, q^0}_{M \text{ times}}, \underbrace{q^K, q^0, \dots, q^0}_{M \text{ times}}, \dots, \underbrace{q^K, q^0, \dots, q^0}_{M \text{ times}} \right).$$

Let $\mathbb{U} = \{1, (M+1)+1, 2(M+1)+1, \dots, (L-1)(M+1)+1\}$ denote the set of promotion periods in $\mathbf{p}^*$. We choose the demand functions f_t to be:

$$f_t(p_t) = \begin{cases} Z/q^K & \text{if } t \in \mathbb{U} \text{ and } p_t = q^K, \\ 1/q^0 & \text{otherwise,} \end{cases}$$

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., $c_t = 0, \forall t = 1, \dots, T$. 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 $t \in \mathbb{U}$.

Step 3: We show that if $\mathbf{p}$ promotes only during periods $t \in \mathbb{U}$, 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 $t \in \mathbb{U}$. Therefore the LP coefficients as defined in (6) are given by:

$$b_t^k = \begin{cases} Z - Y & \text{if } t \in \mathbb{U}, k = K, \\ \leq 0 & \text{otherwise.} \end{cases}$$

Any LP optimal solution selects at most L of γ_t^k , for $k = 1, \dots, K$ 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:

$$(\gamma^{LP})_t^k = \begin{cases} 1 & \text{if } t \in \mathbb{U}, k = K \\ 1 & \text{if } t \notin \mathbb{U}, k = 0 \\ 0 & \text{otherwise} \end{cases}$$

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 $\mathbb{A}$ 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 $t \notin \mathbb{U}$, then we can consider the reduced set of promotions $\mathbb{B} = \mathbb{A} \setminus \{(t, k)\}$. 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 $t + 1, \dots, t + M$. Thus, removing the promotion (t, k) increases the total profit, that is $POP(\gamma(\mathbb{A})) \leq POP(\gamma(\mathbb{B}))$. By applying this procedure repeatedly, one can reach a price vector with only promotions in periods $t \in \mathbb{U}$ 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 $t \in \mathbb{U}$.

Proof of Step 3. Let $\mathbf{p}$ be a price vector that only contains promotions during periods $t \in \mathbb{U}$. Let n be the number of periods t in $\mathbf{p}$ such that $p_t = q^K$ ($n \leq L$ because $\mathbb{U}$ is composed of L periods). Note that all the successive promotions in $\mathbb{U}$ 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. $\square$

2. Upper bound

Proof. Let us denote the bound with n promotions by:

$$\underline{R}_n = \prod_{i=1}^{n-1} g_{i(S+1)}(q^K), \quad (35)$$

when $\underline{R}_0 = 1$ by convention. We can also define the following limit:

$$\underline{R}_\infty = \lim_{n \rightarrow \infty} \underline{R}_n.$$

Note that $g_m(q^K) \leq 1$ so that $\underline{R}_n$ is non-increasing with respect to n . Note also that $g_m(q^K) = 1$ for $m > M$ so that $\underline{R}_{M+1} = \underline{R}_{M+2} = \dots = \underline{R}_\infty$, i.e., the sequence $\underline{R}_n$ converges.

In the case when $S \geq M$, we know from Proposition 1 that the LP approximation is exact. We also know from (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:

$$POP^n = POP(\{q^k\}_{k=0}^K, \{f_t^n\}_{t=1}^{T_n}, \{c_t\}_{t=1}^{T_n}, \{g_m\}_{m=1}^M, L_n, S),$$

where $\{q^k\}_{k=0}^K, \{g_m\}_{m=1}^M, S$ 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:

$$f_t^n(p_t) = \begin{cases} Z/q^K & \text{if } 1 \leq t \leq LM+1 \text{ and } p_t = q^K, \\ 1/q^0 & \text{otherwise.} \end{cases}$$

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:

$$\mathbf{p}^{LP} = \left(q^K, \underbrace{q^0, \dots, q^0}_{S \text{ times}}, q^K, \underbrace{q^0, \dots, q^0}_{S \text{ times}}, \dots, q^K, \underbrace{q^0, \dots, q^0}_{T-(L-1)(S+1)-1 \text{ times}} \right).$$

Step 2: We show that:

$$POP^n(\mathbf{p}_n^{LP}) \leq T - L + Z(\underline{R}_1 + \dots + \underline{R}_n).$$

Step 3: We show the following lower bound for the optimal profit: $POP^n(\mathbf{p}_n^{POP}) \geq nZ$.

Step 4: We finally prove the convergence of the following limit, implying the desired result:

$$\lim_{n \rightarrow \infty} \frac{POP_n(\mathbf{p}_n^{POP})}{POP_n(\mathbf{p}_n^{LP})} = \frac{1}{\underline{R}_\infty}.$$

Proof of Step 1. Based on the above definitions, we have: $POP(\mathbf{p}\{(t, K)\}) = POP(\mathbf{p}^0) + Z - Y$ for $1 \leq t \leq LM + 1$. Therefore, the LP coefficients are given by:

$$b_t^k = \begin{cases} Z - Y & \text{if } 1 \leq t \leq LM + 1, k = K, \\ \leq 0 & \text{otherwise.} \end{cases}$$

Let $\mathbb{U} = \{1, S + 1, 2S + 1, \dots, LS + 1\}$ denote the set of promotion periods in $\mathbf{p}^{LP}$.

Any LP optimal solution selects at most L of γ_t^k , for $k = 1, \dots, K$ 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:

$$(\gamma^{LP})_t^k = \begin{cases} 1 & \text{if } t \in \mathbb{U}, k = K \\ 1 & \text{if } t \notin \mathbb{U}, k = 0 \\ 0 & \text{otherwise} \end{cases}$$

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 $1, 2, \dots, (i - 1)$. In addition, the profit from each non-promotion period is bounded above by 1. We obtain:

$$POP_n(\gamma_n^{LP}) \leq T - L + Z(\underline{R}_1 + \underline{R}_2 + \dots + \underline{R}_n).$$

Proof of Step 3. Consider the following price vector:

$$\mathbf{p} = \left(q^K, \underbrace{q^0, \dots, q^0}_{M \text{ times}}, q^K, \underbrace{q^0, \dots, q^0}_{M \text{ times}}, \dots, q^K, \underbrace{q^0, \dots, q^0}_{M \text{ times}} \right).$$

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 $\frac{1}{R_\infty}$ is both a lower and upper bound of the limit. First, using Theorem 2 for POP^n , we have:

$$\frac{POP^n(\mathbf{p}_n^{POP})}{POP^n(\mathbf{p}_n^{LP})} \leq \frac{1}{R_n}.$$

By taking the limit when $n \rightarrow \infty$ on both sides:

$$\lim_{n \rightarrow \infty} \frac{POP^n(\mathbf{p}_n^{POP})}{POP^n(\mathbf{p}_n^{LP})} \leq \lim_{n \rightarrow \infty} \frac{1}{\underline{R}_n} = \frac{1}{\underline{R}_\infty}.$$

By using Steps 2 and 3, we obtain:

$$\begin{aligned} \lim_{n \rightarrow \infty} \frac{POP^n(\mathbf{p}_n^{POP})}{POP^n(\mathbf{p}_n^{LP})} &\geq \lim_{n \rightarrow \infty} \frac{n \cdot 100nY}{nM + 100nY(\underline{R}_1 + \underline{R}_2 + \dots + \underline{R}_n)} \\ &= \lim_{n \rightarrow \infty} \frac{1}{\frac{M}{100nY} + \frac{\underline{R}_1 + \underline{R}_2 + \dots + \underline{R}_n}{n}} = \frac{1}{\underline{R}_\infty}. \end{aligned}$$

In the last equality, we have used the fact that if $\{a_n\}_{n=1}^\infty$ converges to a finite limit a , then $\{\sum_{i=1}^n a_i/n\}_{n=1}^\infty$ also converges to a .

Appendix D: Proof of Proposition 4

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 (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: $\mathbf{p}_{\{(t_1, k_1), \dots, (t_n, k_n)\}}$. The LP approximation evaluated at this price vector is given by:

$$LP(\mathbf{p}_{\{(t_1, k_1), \dots, (t_n, k_n)\}}) = POP(\mathbf{p}^0) + \sum_{i=1}^n [q^{k_i} POP(\mathbf{p}_{(t_i, k_i)}) - POP(\mathbf{p}^0)].$$

The POP objective using the single promotion (t_i, k_i) is given by:

$$POP(\mathbf{p}_{(t_i, k_i)}) = q^0 f_1(q^0) + \dots + q^0 f_{t_i-1}(q^0) + q^{t_i} f_{t_i}(q^{k_i}) + q^0 [f_{t_i+1}(q^0) + e_{t_i+1}^i] + \dots + q^0 [f_T(q^0) + e_T^i].$$

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: $POP_t = q^0 [f_t(q^0) + e_t^1 + \dots + e_t^i]$. The LP objective at time t is given by:

$$LP_t = q^0 f_t(q^0) + \sum_{j=1}^i \left(q^0 [f_t(q^0) + e_t^j] - q^0 f_t(q^0) \right) = q^0 [f_t(q^0) + e_t^1 + \dots + e_t^i]. \quad (36)$$

As a result, at each time t without a promotion, we have $POP_t = LP_t$ and hence equation (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:

$$POP_t = q^{k_i} [f_{t_i}(q^{k_i}) + e_{t_i}^1 + \dots + e_{t_i}^{i-1}]. \quad (37)$$

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: $q^0 [f_t(q^0) + e_{t_i}^j]$. 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 (36), the LP objective at time t can be written as:

$$LP_t = \sum_{j=1}^{i-1} q^0 e_{t_i}^j + q^{k_i} f_{t_i}(q^{k_i}). \quad (38)$$

By comparing equations (37) and (38), one can see that equation (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 2, we obtain: $0 \geq g_{t_j - t_i}(q^{k_j}) \geq g_{(j-i)(S+1)}(q^K)$. $\square$

Appendix E: Proofs of Tightness for Additive Demand

1. Lower bound

Proof. In the case when $S \geq M$, we know from Proposition 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:

$$POP(\{q^k\}_{k=0}^K, \{f_t\}_{t=1}^T, \{c_t\}_{t=1}^T, \{g_m\}_{m=1}^M, L, S),$$

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:

$$p_t^* = \begin{cases} q^K & t \in \mathbb{U}, \\ q^0 & t \notin \mathbb{U}. \end{cases}$$

Let $\mathbb{U} = \{1, (M + 1) + 1, 2(M + 1) + 1, \dots, (L - 1)(M + 1) + 1\}$ 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:

$$f_t(p_t) = \begin{cases} Z & \text{if } t \in \mathbb{U} \text{ and } p_t = q^K, \\ Y & \text{otherwise.} \end{cases}$$

Note that for any feasible price vector $\mathbf{p}$, the demand at each time is nonnegative. Let us define the costs $c_t = 0, \forall t = 1, \dots, T$. 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 definition, we have: $POP(\mathbf{p}\{(t, K)\}) - POP(\mathbf{p}^0) = q^K Z - q^0 Y - q^0 Y$ for $t \in \mathbb{U}$. 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 $t+1, \dots, t+M$ of $POP(\mathbf{p}\{(t, K)\})$ due to the promotion in period t . Therefore, the LP coefficients as defined in (6) are:

$$b_t^k = \begin{cases} q^K Z - 2q^0 Y \geq 0 & \text{if } t \in \mathbb{U}, k = K \\ \leq 0 & \text{otherwise} \end{cases}$$

The LP optimal solution selects at most L of γ_t^k , for $k = 1, \dots, K$ to be 1. Consequently, the optimal LP objective is bounded above by $Tq^0 Y + L(q^K Z - 2q^0 Y)$. In fact, the following γ^* corresponding to $\mathbf{p}^*$ achieves this bound and is therefore optimal:

$$(\gamma^*)_t^k = \begin{cases} 1 & \text{if } t \in \mathbb{U}, k = K \\ 1 & \text{if } t \notin \mathbb{U}, k = 0 \\ 0 & \text{otherwise} \end{cases}$$

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 $\mathbb{U}$ and the second term is the profit from the non-promotion periods $\mathbb{T} \setminus \mathbb{U}$ before promotions. Finally, the third term represents the reduction in profit during the non-promotion periods due to the promotions in $\mathbb{U}$.

Let POP_t be the $POP(\mathbf{p})$ profit at period t . If we promote at time $t \in \mathbb{U}$ 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 $t \in \mathbb{U}$. 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 $\mathbb{U}$, whereas the second term comes from the non-promotion periods. One can see that:

$$\begin{aligned} POP(\mathbf{p}^*) - POP(\mathbf{p}) &= Lq^K Z + (T - L)q^0 Y - Lq^0 Y - [(L - 1)q^K Z + (T - L + 1)q^0 Y] \\ &= q^K Z - (L + 1)q^0 Y \geq 0, \end{aligned}$$

from the definition of Z . Therefore, $POP(\mathbf{p}^*) \geq POP(\mathbf{p})$ as desired. $\square$

2. Upper bound

Proof. In the case when $S \geq M$, we know from Proposition 1 that an optimal solution of the LP is also an optimal solution of the POP. We also know from equation (22) that $\bar{R} = 0$. Thus, the result holds in this case.

In the case when $S < M$, we will construct a POP problem:

$$POP(\{q^k\}_{k=0}^K, \{f_t\}_{t=1}^T, \{c_t\}_{t=1}^T, \{g_m\}_{m=1}^M, L, S),$$

an optimal LP price vector $\mathbf{p}^{LP}$, and an optimal POP price vector $\mathbf{p}^{POP}$, such that $POP(\mathbf{p}^{POP}) = POP(\mathbf{p}^{LP}) + \bar{R}$. Let $T = (M+1)L$. Let us define $Y = \sum_{i=1}^M |g_i(q^K)|$, $Z = (L+1)q^0Y/q^K$ 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:

$$\mathbf{p}^{LP} = \left(q^K, \underbrace{q^0, \dots, q^0}_{M \text{ times}}, q^K, \underbrace{q^0, \dots, q^0}_{M \text{ times}}, \dots, q^K, \underbrace{q^0, \dots, q^0}_{M \text{ times}} \right).$$

Step 2: We show that the following price vector is an optimal POP solution:

$$\mathbf{p}^{POP} = \left(q^K, \underbrace{q^0, \dots, q^0}_{S \text{ times}}, q^K, \underbrace{q^0, \dots, q^0}_{S \text{ times}}, \dots, q^K, \underbrace{q^0, \dots, q^0}_{T-(L-1)(S+1)-1 \text{ times}} \right).$$

Step 3: We show that $POP(\mathbf{p}^{POP}) = POP(\mathbf{p}^{LP}) + \bar{R}$ 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 $t \in \mathbb{U}$. 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 $t+1, \dots, t+M$ of $POP(\mathbf{p}\{(t, K)\})$ due to the promotion in period t . Therefore, the LP coefficients as defined in (6) are:

$$b_t^k = \begin{cases} q^K Z - 2q^0 Y \geq 0 & \text{if } t \in \mathbb{U}, k = K \\ \leq 0 & \text{otherwise} \end{cases}$$

The LP optimal solution selects at most L of γ_t^k , for $k = 1, \dots, K$ to be 1. Consequently, the optimal LP objective is bounded above by $Tq^0Y + L(q^K Z - 2q^0Y)$. In fact, the following γ^{LP} corresponding to $\mathbf{p}^{LP}$ achieves this bound and is therefore optimal:

$$(\gamma^{LP})_t^k = \begin{cases} 1 & \text{if } t \in \mathbb{U}, k = K \\ 1 & \text{if } t \notin \mathbb{U}, k = 0 \\ 0 & \text{otherwise} \end{cases}$$

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:

$$POP(\mathbf{p}^{LP}) = LP(\mathbf{p}^{LP}). \quad (39)$$

Proof of Step 2. By using Proposition 4, 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 $\bar{R}$ from (22), one can see that $ER(\mathbf{p}) \leq \bar{R}$ for all feasible $\mathbf{p}$. In other words, $\bar{R}$ corresponds to the largest possible error term. In addition, we have in this case: $ER(\mathbf{p}^{POP}) = \bar{R}$ 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:

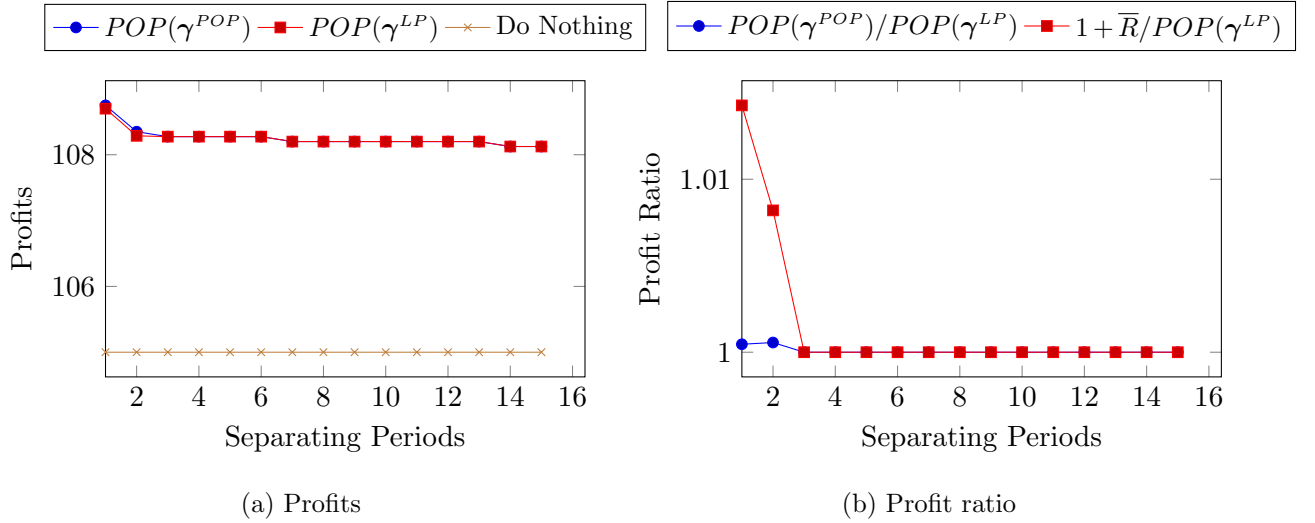
$$POP(\mathbf{p}^{POP}) = LP(\mathbf{p}^{POP}) + \bar{R}. \quad (40)$$

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 (39) and (40) gives us the desired result. $\square$

Appendix F: Additive demand: illustrating the bounds

In what follows, we test numerically the upper bound for the additive model from Section 6.2 by varying the different model parameters. In Figures 11, 12 and 13, the demand model is given by: $d_t(\mathbf{p}) = 30 - 50p_t + 15p_{t-1} + 10p_{t-2} + 5p_{t-3}$.

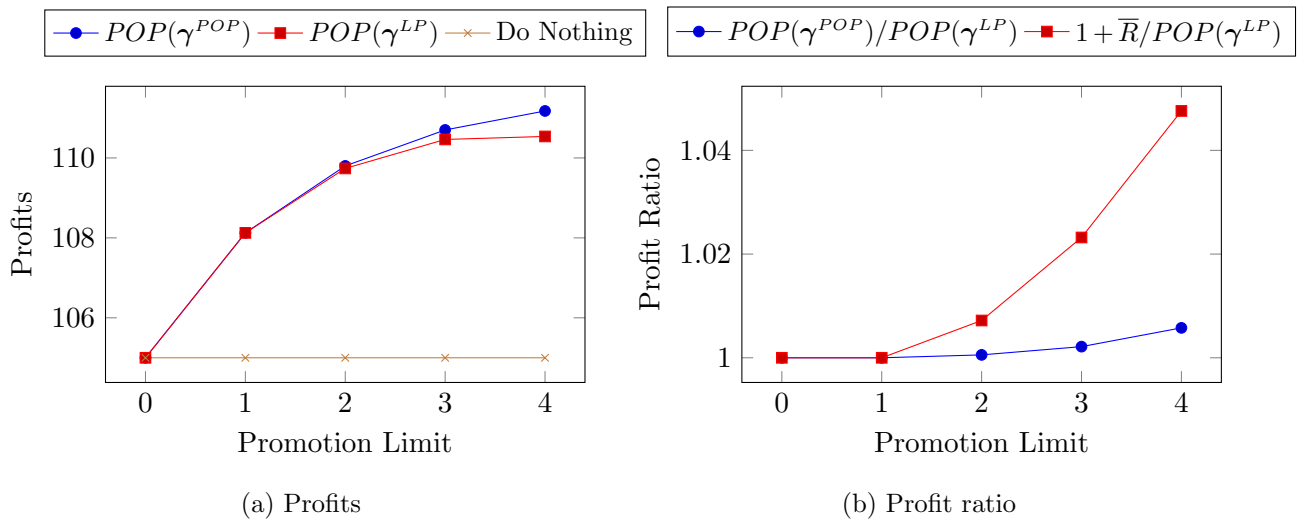
Figure 11 Effect of varying the separating periods S



Note. Example parameters: 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 11, we vary the number of separating periods S . We make the following observations: *a)* As one would expect from Proposition 1, the LP approximation coincides with the optimal POP solution for $S \geq M = 4$. *b)* As expected, as S increases, the upper bound $1 + \bar{R}/POP(\gamma^{LP})$ decreases. Indeed, the larger is S , the more separated promotions are and as a result, it 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.

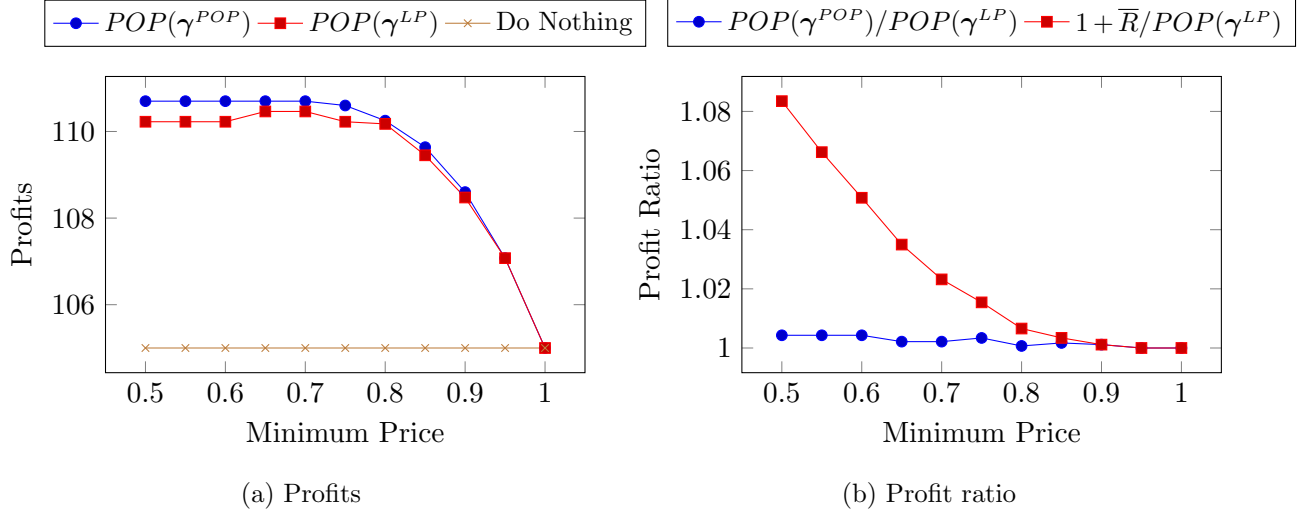
Figure 12 Effect of varying the promotion limit L



Note. Example parameters: $S = 0$, $\mathcal{Q} = \{1, 0.95, 0.90, 0.85, 0.80, 0.75, 0.70\}$.

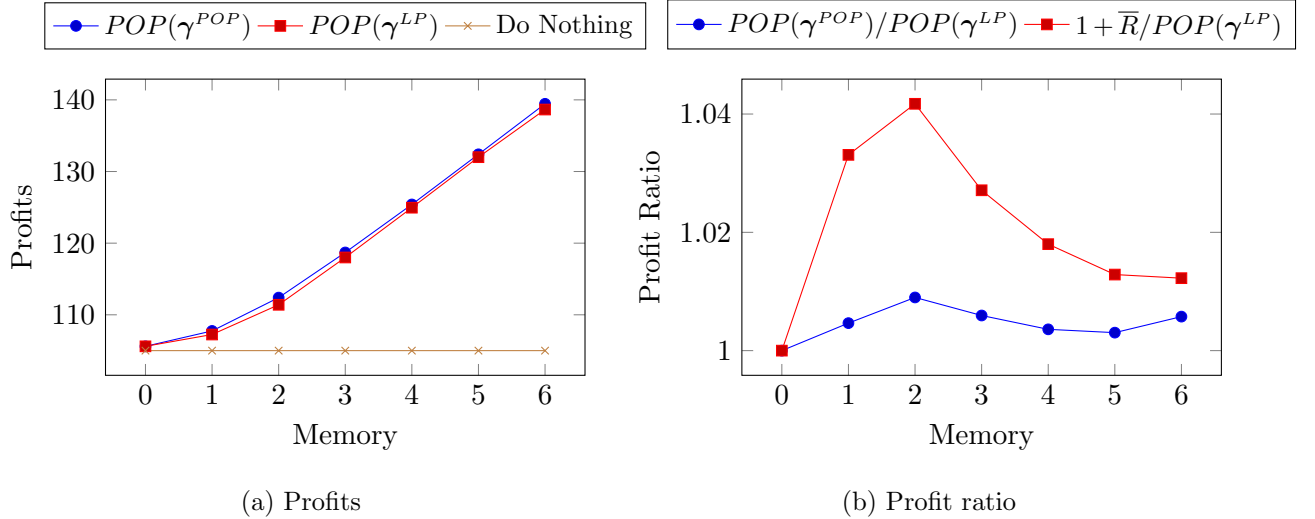
Dependence on the number of promotions allowed: In Figure 12, we vary the number of promotions allowed L . We make the following observations: *a)* As one would expect from Proposition 1, the LP approximation coincides with the optimal POP solution for $L = 1$. *b)* For $L \leq 6$ (recall that $T = 13$), 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 $\bar{R}$ in Theorem 3. Unlike the multiplicative case for which $\underline{R}$ was asymptotically converging as L increases; in the additive case, $\bar{R}$ can grow to infinity as L increases.

Dependence on the minimal price of the price ladder: In Figure 13, 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

Figure 13 Effect of varying the minimum price q^K 

Note. Example parameters: $L = 3, S = 0$.

most 2.5%. From the definition of $\bar{R}$ in Theorem 3, one can see that the additive contribution $\bar{R}$ increases as q^K decreases.

Figure 14 Effect of varying the memory M 

Note. Example parameters: $d_t(\mathbf{p}) = 30 - (20 + 20M)p_t + 20p_{t-1} + 20p_{t-2} + \dots + 20p_{t-M}$, $L = 3, S = 0$.

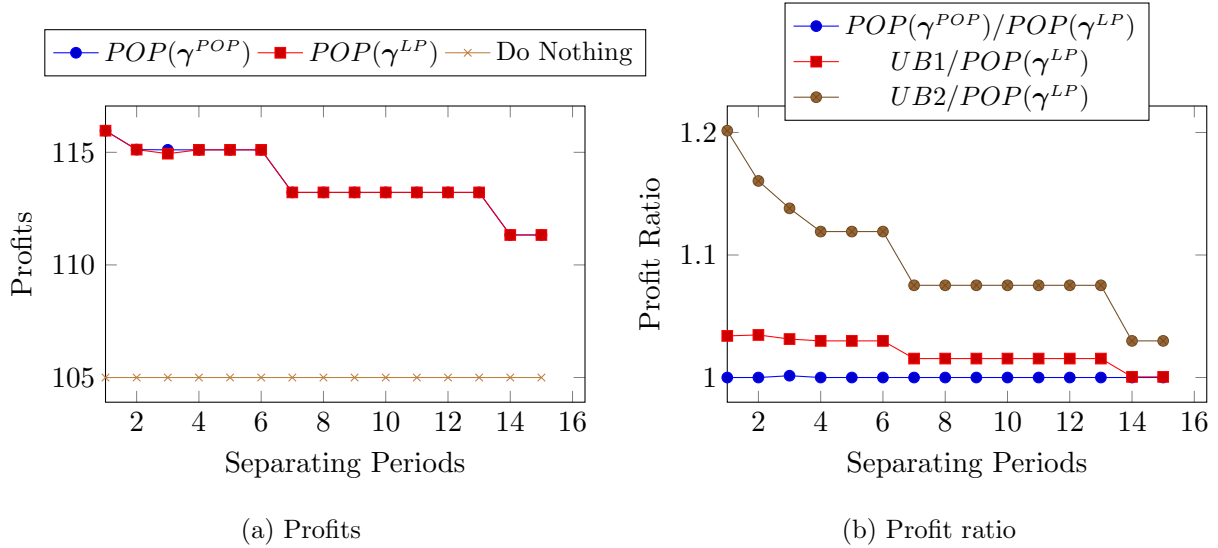
Dependence on the length of the memory: In Figure 14, we vary the memory parameter M . Note that in this example, we have chosen equal coefficients for $g_1, g_2, \dots, g_M$, as a “worst case” so that past prices have a uniformly strong effect on current demand. We make the following observations: a) As one would expect from Proposition 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 $\bar{R}$ in Theorem 3, one can see that $\bar{R}$ 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 $\bar{R}$. Indeed, from the definition of $\bar{R}$, 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 unit will not increase $\bar{R}$. In this case, the terms $g_{(j-i)(S+1)}(q^K)$ do not change.

Appendix G: Unified demand: illustrating the bounds

In what follows, we test numerically the upper bounds for the unified model from Section 6.3 by varying the different parameters of the model. In Figures 15, 16 and 17, 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 (31) and (30) respectively as well as the performance of the LP approximation for the above unified demand model.

Figure 15 Effect of varying the separating parameter S

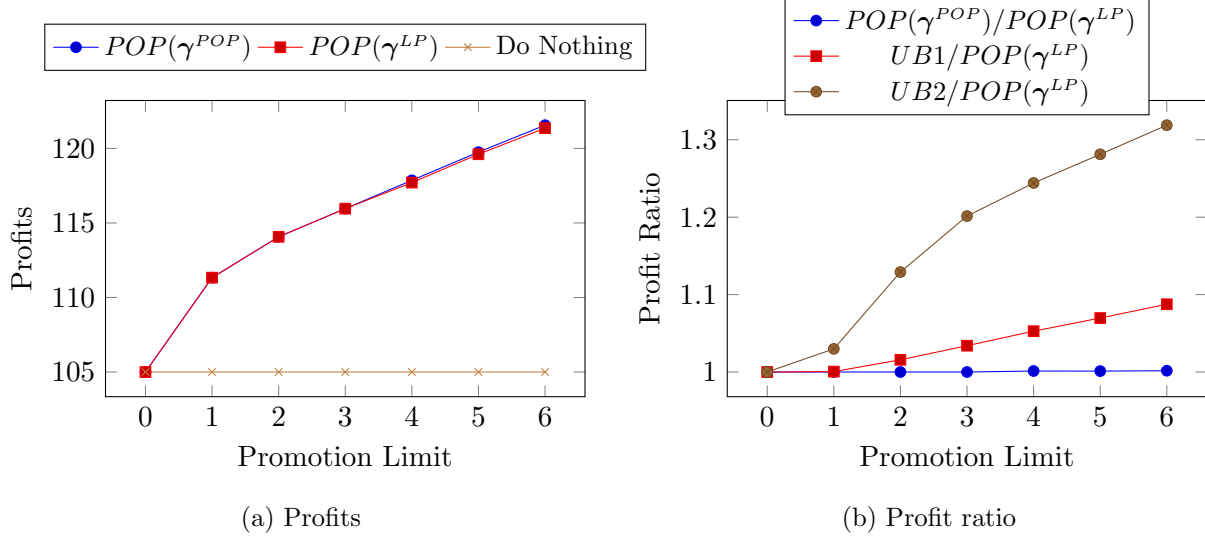


Note. Example parameters: $L = 3$, $\mathcal{Q} = \{1, 0.95, 0.90, 0.85, 0.80, 0.75, 0.70\}$.

Dependence on separating periods: In Figure 15, we vary the number of separating periods S . We make the following observations: *a)* As one would expect from Proposition 1, the LP approximation coincides with the optimal POP solution for $S \geq M = 4$. *b)* As expected, as S increases, the upper bound $1 + \bar{R}/POP(\gamma^{LP})$ 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.

Figure 16 Effect of varying the number of promotions allowed L

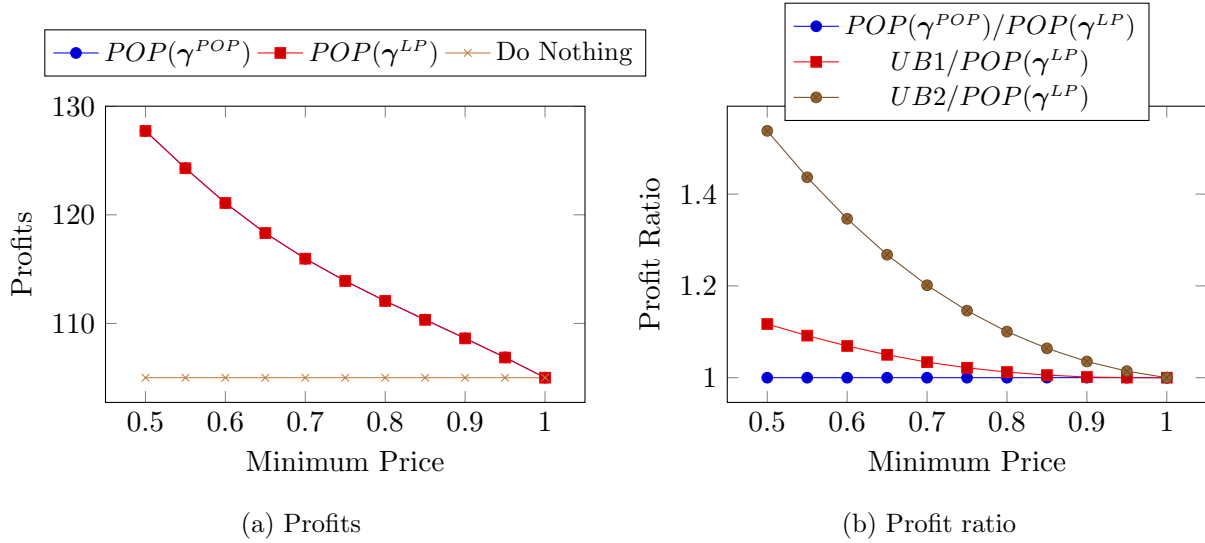


Note. Example parameters: $S = 1$, $\mathcal{Q} = \{1, 0.95, 0.90, 0.85, 0.80, 0.75, 0.70\}$.

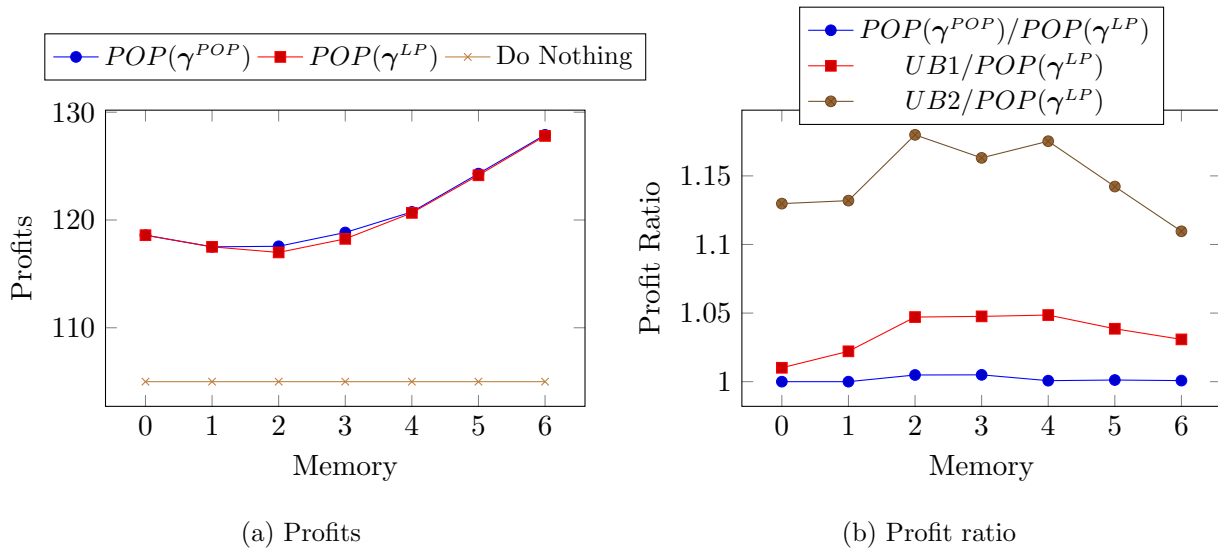
Dependence on the number of promotions allowed: In Figure 16, we vary the number of promotions allowed L . We make the following observations: a) As one would expect from Proposition 1, the LP approximation coincides with the optimal POP solution for $L = 1$. b) For $L \leq 6$ (recall that $T = 13$), 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 $\bar{R}$ in (22) is increasing with L . Unlike the multiplicative case for which $\bar{R}$ was asymptotically converging as L increases; in the additive case, $\bar{R}$ 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 .

Dependence on the minimal price of the price ladder: In Figure 17, 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 $\bar{R}$ in Theorem 3, one can see that $\bar{R}$ increases as q^K decreases.

Dependence on the length of the memory: In Figure 18, we vary the memory parameter M . Note that in this example, we have chosen equal coefficients for $g_1, g_2, \dots, g_M$, 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 Proposition 1, the LP approximation coincides with the

Figure 17 Effect of varying the minimum price q^K 

Note. Example parameters: $L = 3, S = 1$.

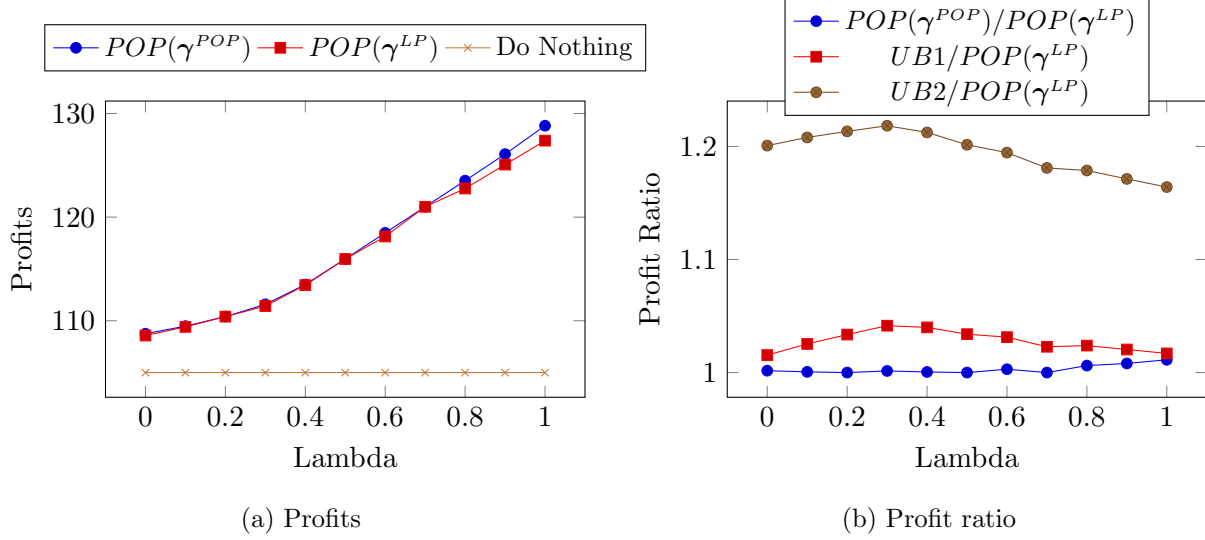
Figure 18 Effect of varying the memory parameter M 

Note. Example parameters: $d_t(\mathbf{p}) = 0.5(10p_t^{-4}p_{t-1}^{0.2}p_{t-2}^{0.2}\dots p_{t-M}^{0.2}) + 0.5[30 - (20 + 20M)p_t + 20p_{t-1} + 20p_{t-2} + \dots + 20p_{t-M}]$, $L = 3, S = 1$.

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 $\bar{R}$ in (22), one can see that the additive contribution $\bar{R}$ 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 $\bar{R}$. Indeed, from the definition of $\bar{R}$, 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 $\bar{R}$. In this case, the terms $g_{(j-i)(S+1)}(q^K)$ do not change.

Figure 19 Effect of varying the parameter λ



Note. Example parameters: $d^{\text{mult}}(p) = 10p_t^{-4}p_{t-1}^{0.2}\cdots p_{t-M}^{0.2}$, $d^{\text{add}}(p) = 30 - 20p_t + 2p_{t-1} + \cdots + 2p_{t-M}$, $d_t(\mathbf{p}) = \lambda d^{\text{mult}}(p) + (1 - \lambda)d^{\text{add}}(p)$, $L = 3, S = 1$.

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 19, 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 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, $\lambda = 0.3$).