informs
http://pubsonline.informs.org/journal/opre/

OPERATIONS RESEARCH
Vol. 65, No. 2, March–April 2017, pp. 446–468
ISSN 0030-364X (print), ISSN 1526-5463 (online)

The Impact of Linear Optimization on Promotion Planning

Maxime C. Cohen,[a] Ngai-Hang Zachary Leung,[b] Kiran Panchamgam,[c] Georgia Perakis,[d] Anthony Smith[c]

[a] Stern School of Business, New York University, New York, New York 10012; [b] College of Business, City University of Hong Kong, Kowloon, Hong Kong; [c] Oracle Retail Global Business Unit (RGBU), Burlington, Massachusetts 01803; [d] Sloan School of Management, Massachusetts Institute of Technology, Cambridge, Massachusetts 02139

Contact: maxime.cohen@stern.nyu.edu (MCC); znhleung@cityu.edu.hk (N-HZL); kiran.panchamgam@oracle.com (KP); georgiap@mit.edu (GP); anthony.x.smith@oracle.com (AS)

Received: January 20, 2014
Revised: August 7, 2015; June 27, 2015
Accepted: September 23, 2016
Published Online in Articles in Advance: January 19, 2017

Subject Classifications: programming: integer, linear/applications, nonlinear/applications
Area of Review: Optimization

https://doi.org/10.1287/opre.2016.1573

Abstract. Sales promotions are important in the fast-moving consumer goods (FMCG) industry due to the significant spending on promotions and the fact that a large proportion of FMCG products are sold on promotion. This paper considers the problem of planning sales promotions for an FMCG product in a grocery retail setting. The category manager has to solve the promotion optimization problem (POP) for each product, i.e., how to select a posted price for each period in a finite horizon so as to maximize the retailer's profit. Through our collaboration with Oracle Retail, we developed an optimization formulation for the POP that can be used by category managers in a grocery environment. Our formulation incorporates business rules that are relevant, in practice. We propose general classes of demand functions (including multiplicative and additive), which incorporate the post-promotion dip effect, and can be estimated from sales data. In general, the POP formulation has a nonlinear objective and is NP-hard. We then propose a linear integer programming (IP) approximation of the POP. We show that the IP has an integral feasible region, and hence can be solved efficiently as a linear program (LP). We develop performance guarantees for the profit of the LP solution relative to the optimal profit. Using sales data from a grocery retailer, we first show that our demand models can be estimated with high accuracy, and then demonstrate that using the LP promotion schedule could potentially increase the profit by 3%, with a potential profit increase of 5% if some business constraints were to be relaxed.

Funding: This work was also partly supported by the National Science Foundation [Grant CMMI-1162034].
Supplemental Material: The online appendix is available at https://doi.org/10.1287/opre.2016.1573.

Keywords: promotion optimization • dynamic pricing • integer programming • retail operations

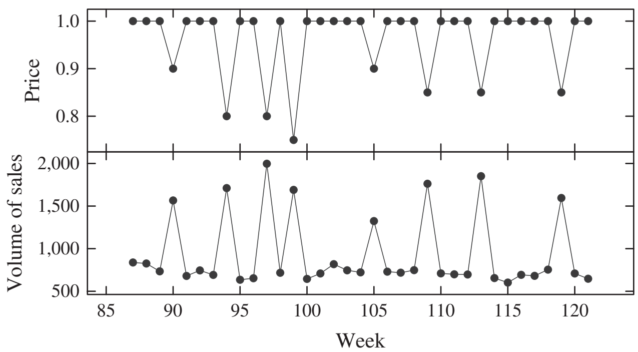
1. Introduction

Fast-moving consumer goods (FMCG) or consumer packaged goods (CPG) are products, which are consumed quickly. Examples of FCMG include processed foods and drinks (e.g., canned food, soft drinks, salty snacks, candy, and chocolate) as well as household products (e.g., toiletries, laundry detergent). The FMCG industry includes some of the most widely recognized companies in the world, such as Coca-Cola, Pepsico, Kraft, Nestle, Proctor & Gamble, Johnson & Johnson, and Unilever. Typically, consumers do not purchase FMCG products directly from manufacturers, but rather from retailers (supermarkets and grocery stores).

It is common for manufacturers and retailers in the FMCG industry to use promotions to entice consumers to purchase their products. The reasons why promotions are used include: increasing sales and traffic, introducing new items, bolstering customer loyalty, competitive retaliation, and price discrimination. The amount of money spent on promotions is significant—it is estimated that FMCG manufacturers spend about $1 trillion annually on promotions (Nielsen 2014). In addition, promotions play an important role in the FMCG industry as a significant proportion of sales are made on promotion. For example, Nielsen (Nielsen 2014) found that 12%–25% of supermarket sales in five European countries (Great Britain, Spain, Italy, Germany, and France) were made on promotion (Gedenk et al. 2006). Not all retailers use promotions—some retailers (e.g., Walmart) employ an *everyday low price* policy. In this paper, we focus on retailers that use price promotions.

A promotion tactic that is commonly used by retailers is *temporary price reductions*. We illustrate the effectiveness of temporary price reductions in boosting sales using real data. In Figure 1, we plot the weekly (normalized) prices and sales for a particular brand of ground coffee in a supermarket over 35 weeks. We observe that this brand was promoted during 8 out of 35 weeks (i.e., 23% of the time); and that promotional sales accounted for 41% of the total sales volume. In this paper, we focus on temporary price reductions by grocery retailers that we simply refer to as promotions.

Figure 1. Prices and sales for a particular brand of coffee at a supermarket over a span of 35 weeks



Promotions can have a significant impact on a retailer's profitability. Using a demand model estimated from sales data (see Section 7.3 for details), we estimate that the promotions set by 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 Fund reports that the average profit margin for the supermarket industry was 1.9% in 2010. According to an analysis performed with Yahoo! Finance data, the average net profit margin for publicly traded U.S.-based grocery stores for 2012 is close to 2010's 1.9% average. As a result, this suggests that promotions can have a significant impact on the retailer's profits. Furthermore, this 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?

Given the importance of promotions in the grocery industry, it is not surprising that supermarkets pay great attention to design promotion schedules. The promotion planning process is complex and challenging for multiple reasons. First, demand is affected by a *post-promotion dip effect*, i.e., for certain categories of products, promotions lead to reduced future demand. 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 (for more details, see Section 3.2). Finally, the problem is difficult even for a single retail store because of its large scale. For instance, a typical supermarket in the United States carries about 40,000 SKUs, with approximately 2,000 SKUs on promotion at any point in time, which leads to a very large number of decisions variables.

Despite the complexity of the promotion planning process, it is still to this day, performed manually in most supermarkets. 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 retailers.

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, because our formulation can be solved quickly, a manager can test various what-if scenarios to study the robustness of the solution.

Our proposed POP formulation is a nonlinear integer programming (IP). For general demand functions, the formulation is NP-hard (see Cohen et al. 2016). An important business rule for retailers, in practice, is that the price of the product must be selected from a price ladder, i.e., a discrete set of permissible prices. In addition, due to the post-promotion dip effect, for general demand functions, the objective is neither concave nor convex. We therefore 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 a general demand. In addition, we derive 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, the profits are supermodular in promotions. In this context, submodular (supermodular) means that the marginal effect on an additional promotion has a smaller (larger) impact when we already have many promotions in the selling season. 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.

The impact of our models can be also significant for supermarkets, in practice. One of the goals of this research has been 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 midmarket (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 is the large-scale nature of this industry. For example, a typical Tier 1 retailer has approximately 1,000 stores, with 200 categories each containing 50–600 items. An important criterion for our models to be adopted by grocery retailers, is that the software tool needs to run in a few seconds up to a minute. This motivated us to reformulate our model as an LP.

Preliminary tests using supermarket sales 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 coauthors 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 not computationally tractable, in general. 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 a software tool embedding the model and algorithms 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 the IP can be solved efficiently as an LP.
- *We introduce general classes of demand functions that model the post-promotion dip effect*. An important feature of the application domain is the post-promotion dip in demand observed, in practice. 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 models commonly used in the literature, provide modeling flexibility and can be estimated from data.
- *We develop tight 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 computationally 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.
- *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 several product categories. We apply our analysis to a few selected categories (ground coffee, tea, chocolate, and yogurt). We first estimate the 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.
- *We demonstrate that our results are robust with respect to demand uncertainty*. We propose a way to address the case where the estimated demand parameters are uncertain. We then validate the robustness of our solution using actual data. In particular, extensive testing suggests that the profit gain dominates the forecasting error.

2. Literature Review

Our work is related to four streams of literature: optimization, marketing, dynamic pricing, and retail operations. We formulate the POP for a single item as a nonlinear mixed-integer program (NMIP). To give users flexibility in the choice of demand functions, our POP formulation imposes very mild assumptions on the demand. Due to the general classes of demand functions, we consider the objective is typically nonconcave. 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 (see, e.g., Bertsimas and Shioda 2009, 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 the 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 IP reformulation techniques to potentially address the nonconvexities. 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.

Our work is related to linearization techniques for nonlinear IP problems. One common procedure in this field is to add additional constraints, and possibly introduce new variables, to produce tight (usually linear) relaxations. It is, of course, necessary to prove that the new problem is equivalent to the initial one. We briefly describe two important works in this area. In Adams et al. (2004), the authors present a strategy for finding a tighter linear relaxation of a mixed 0-1 quadratic program. In Chaovalitwongse et al. (2004), the authors propose a new linearization method for quadratic 0-1 programming problems with linear and quadratic constraints, which requires only $O(kn)$ additional continuous variables where k denotes the number of quadratic constraints.

In this paper, we formulate the POP as a nonlinear binary IP problem. Due to the general nature of the demand functions, to the best of our knowledge, none of the existing approaches in the literature apply directly to our formulation. Instead of solving the nonlinear problem exactly, we consider an approximation based on linearizing the nonlinear objective function. The linearization uses the sum of the marginal contributions of a single promotion, and can be viewed as a first-order Taylor approximation of the total profits around the regular prices. By showing that the constraint matrix of the integer program is totally unimodular, solving the linear relaxation yields the optimal IP solution. By exploiting the structure of the demand functions, we derive provable performance guarantees for our solution method.

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). 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 objectives, maximizing a submodular function 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 nonmonotone submodular function subject to no constraints admits a $1/2$ approximation algorithm (see, for example, Buchbinder et al. 2012, 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 any structure on the objective. This LP approximation also provides a guarantee relative to the optimal profit for two 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 well studied in marketing (see Blattberg and Neslin 1990, and the references therein). The marketing community has observed that for many FMCG products, temporary price reductions lead to a future demand reduction, a phenomenon that is referred to as the *post-promotion dip* effect. Marketing researchers typically focus on developing and estimating demand models, e.g., linear regression or choice models, to derive managerial insights on promotions (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. One of the methods used in the marketing literature to capture the post-promotion dip effect is to use a demand function that depends not only on the current price, but also on the prices at the most recent periods (Mela et al. 1998, Heerde et al. 2000, Macé and Neslin 2004, Ailawadi et al. 2007).

Our work is also related to the field of dynamic pricing (see, e.g., Talluri and van Ryzin 2005 and the references therein). More specifically, in the operations management literature, researchers study sales promotions from the angle of how to optimize the dynamic pricing of a product given that consumer behavior leads to post-promotion dips in demand. One approach is to build a game-theoretic model of consumers. In Assunção and Meyer (1993), the authors consider the problem faced by a rational consumer regarding optimal purchasing and consumption of a storable good. In their model, the price in the next period is assumed to be random (drawn from a stationary distribution of prices conditional on the last observed price). In addition, the authors assume that the seller's pricing policy is exogenous and random. In Su (2010), the author develops a model with multiple consumer types who may differ in their holding costs, consumption rates, and fixed shopping costs. The author could solve the dynamic pricing model for the rational expectation equilibrium, and draws several managerial insights. An alternative approach used in the dynamic pricing literature is to model consumers using a reference price model (Kopalle et al. 1996, Fibich et al. 2003, Popescu and Wu 2007, Chen et al. 2016). The reference price model posits that consumers form an internal reference price for the product based on past price observations. When the consumer observes the current price, she compares it to the reference price as a benchmark. Prices above the reference price are perceived to be "high," which leads to lower demand, whereas prices below the reference price are perceived to be "low," leading to an increase in demand. The papers by Kopalle et al. (1996), Fibich et al. (2003), Popescu and Wu (2007) study an infinite-horizon dynamic pricing problem with a reference price model. In Chen et al. (2016), the authors analyze a periodic review stochastic inventory model in which pricing and inventory decisions are made simultaneously. Our paper differs from the models in the dynamic pricing literature in that our dynamic pricing problem includes business rules that are relevant, in practice.

In Ahn et al. (2007), the authors propose a demand model in which a proportion of customers will wait k periods after and will purchase once the posted price of the product falls below their willingness to pay. Under this assumption, their demand model depicts the post-promotion dip effect.

In this paper, we propose two general classes of multiplicative and additive demand models. The multiplicative model is a generalization of the linear regression model with lagged variables used by Heerde et al. (2000), Macé and Neslin (2004). In addition, the demand model we propose can closely approximate the reference price model used in Kopalle et al. (1996), Fibich et al. (2003), Popescu and Wu (2007); as well as the demand model in Ahn et al. (2007). Finally, our work is related to the field of retail operations, and more specifically, pricing problems under business rules. One of the constraints considered in our paper imposes the prices to lie in a discrete set. Zhao and Zheng (2000) consider a dynamic pricing problem for a fixed inventory perishable product sold over a finite- (continuous) time horizon. For the case of a discrete price set, the authors solve the continuous time dynamic program by applying a discretization approach and a backward recursion. The computational complexity of their approach grows linearly with the number of discrete-time intervals. Our approach is different in nature and is based on an LP approximation that yields a complexity polynomial in the number of time periods. Subramanian and Sherali (2010) study a pricing problem for grocery retailers, where prices are subject to interitem constraints. They propose a linearization technique to solve the problem. Caro and Gallien (2012) study a markdown pricing problem for a fashion retailer, for which the prices are constrained to be nonincreasing, and some set of items 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, the assumptions, the business rules, and we formulate the POP. In Section 4, we present an approximate formulation based on linearizing the objective, which gives rise to a linear IP. We then show that the IP can, in fact, be solved as an LP. In Section 5, we consider multiplicative and additive demand models and derive tight bounds on the LP approximation relative to the optimal solution. In Section 6, we consider an extension of our approach for uncertain demand. Section 7 presents computational results using real data. Finally, we present our conclusions. Several of the proofs of the different propositions and theorems are relegated to the appendix.

3. Model, Assumptions, and Problem Formulation

In this section, we present a mathematical model of the POP for an FMCG product. One of our primary goals is to incorporate problem features that are relevant, in practice. The model was developed through a collaboration with our coauthors working at Oracle Retail, and thus we have benefited from the expertise of Oracle executives as well as retailers.

The manager of an FMCG category in a grocery retailer faces the POP: for a given product, how to select a posted price for each period in a finite sales horizon so as to maximize the retailer's profit. In the following, we describe the assumptions underlying our formulation (see Section 3.1) as well as the business rules (see Section 3.2). Finally, we present a mathematical formulation of the POP in Section 3.3.

3.1. Assumptions

In this paper, we focus on a single-item model of the POP.

Assumption 1 (Cost of Inventory). *At each period t, the retailer orders inventory from the supplier at a unit cost c_t.*

The above assumption holds under the conventional wholesale price contract, which is frequently used, in practice, and in the academic literature (see, e.g., Cachon and Lariviere 2005, Porteus 1990).

Assumption 2 (Demand is a Deterministic Function of Prices). *The demand in period t is a deterministic function of the prices chosen by the retailer $(p_1, p_2, \ldots, p_T)$.*

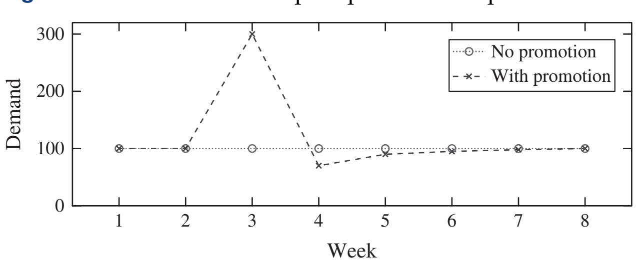
Here, T denotes the length of the horizon. This assumption is reasonable for FMCG products because the prices can be used to accurately forecast demand. This assumption is also supported by our experiments, which show that a regression model using past prices is able to predict future demand with a low forecast error (see estimation results in Section 7 and Figure 4). Since the estimated deterministic demand functions seem to accurately model actual demand, for this application, we can use them as input to the optimization model without taking into account demand uncertainty. We also propose an extension of our approach to relax the deterministic demand assumption, and address the case where demand is uncertain (see Section 6).

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 compute the optimal prices using our model. The demand models we consider are commonly used by practitioners and the academic literature (see Heerde et al. 2000, Macé and Neslin 2004, Fibich et al. 2003).

Post-Promotion Dip in Demand. As we mentioned, the demand of an FMCG product often has the *post-promotion dip* property. In particular, promoting the product in period $u < t$, may reduce the demand in period t (relative to the demand value if the product was not promoted at time u). This is illustrated in Figure 2. We model the post-promotion dip property by assuming that the demand is a function of the current price and the prices in the M most recent periods:

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

Figure 2. Illustration of the post-promotion dip effect



Notes. A promotion in week 3 boosts the demand in week 3, but decreases the demand in the following weeks. Demand then gradually recovers up to the no-promotion level.

It has been recognized in the marketing literature that in some retail settings, following a promotion, there is a decline in sales relative to what sales would have been in the absence of a promotion. This is referred to as a *post-promotion dip in sales* (see, e.g., Macé and Neslin 2004).

There are multiple possible explanations regarding the post-promotion dip effect. One explanation is that consumers respond to promotions by purchasing larger quantities, which they stockpile for future consumption. Another explanation is related to the reference price effect. Very often, consumers form a *reference price* for the product, which is a weighted average of the recent observed prices. The current demand is then affected not only by the current price, but also by the reference price, through the psychological effect of feeling a gain or a loss. If the current price is higher than the reference price, consumers perceive the current price as a *loss*, which decreases demand; conversely, if the current price is lower than the reference price, consumers perceive the current price as a *gain*, which increases demand. The reference price model has been used in dynamic pricing problems, see, e.g., Kopalle et al. (1996), Fibich et al. (2003), Popescu and Wu (2007), Chen et al. (2016).

Alternatively, instead of looking at individual price effects on demand, one may look at the difference between the effect of the current price p_t relative to some weighted average of past prices (e.g., p_{t-1} and p_{t-2} when $M = 2$). Equivalently, the consumers use some sort of weighted average as a reference point. Then, the two following effects may be observed.

(a) Comparison effect: Consider the case where the current price p_t is higher relative to the past prices p_{t-1} and p_{t-2}. Then, compared to the possibility that the consumers could have purchased the product at lower prices in the past, buying the product at this time (with the higher price p_t) feels like a loss to consumers. The greater the difference between p_t and p_{t-1} (or p_{t-2}) is, the greater the sensation of loss the consumers feel for buying now. Such a comparison induces the consumers to become less willing to buy at p_t, hence decreasing the demand.

(b) Attachment effect: Consider now the case where the current price p_t is smaller relative to the past prices p_{t-1} and p_{t-2}. As the consumers were expecting to pay a higher price, purchasing the product at the lower price p_t feels like a gain. In addition, the greater the difference between p_{t-1} (or p_{t-2}) and p_t is, the greater the sensation of gain the consumers feel. Therefore, the consumers become more attached to the idea of buying at this period due to this gain feeling. This attachment effect increases the consumers' willingness as well as the demand.

Assumption 3 (Sufficient Inventory). *The retailer has sufficient inventory to meet demand in each period, i.e., sales is equal to demand.*

Remark 1. The assumption that the retailer carries enough inventory to meet demand does not apply to all products and all retail settings. For example, it is common practice in the fashion industry (e.g., Talbots or Zara) to intentionally produce limited amounts of inventory. By doing so, a retailer sends a signal to consumers that they should buy the product now at the regular price. If the consumers decide to wait until the clearance season, there is a risk that the product would be sold out. Consumers also expect products to stock out as the seasons change (e.g., spring to summer).

Unlike fashion items, which go out of season, FMCG products such as ground coffee or soft drinks are typically available all year round. These products typically have shelf lives greater than six months, and customers have been conditioned to expect that these products would always be in stock at retail stores. Since these products are easy to store and have a high degree of availability, FMCG retailers typically do not use the risk of stock out to incentivize consumers to buy now.

In addition, as we will show in our computational experiments, the demand forecast accuracy is high and the out-of-sample (OOS) metrics are very good (the OOS R^2 and mean absolute percentage error (MAPE)). In the data we have, we actually observed that the inventory was not issue and saw very few events of stock-outs over a two-year period. This can be justified by the fact that supermarkets have a long experience with inventory decisions and accumulated large data sets allowing them to develop sophisticated forecasting demand tools to support capacity and ordering decisions. Many such models were developed in the last two decades (e.g., Cooper et al. 1999, Van Donselaar et al. 2006). Finally, it seems reasonable that consumers buying behavior is easier to predict for grocery products such as ground coffee, relative to fashion items, which are unique and have short product cycles. Indeed, fashion involves an impulsive and occasional purchasing behavior (and hence can be harder to predict), whereas grocery items are more routinely-based purchases. Finally, grocery retailers are aware of the negative effects of stocking out of promoted products (see, e.g., Corsten and Gruen 2004, Campo et al. 2000). For all of the above reasons, the statement in Assumption 3 that the retailer carries sufficient inventory to meet demand in each period is reasonable in the context of FMCG products.

3.2. Business Rules

Business Rule 1 (Discrete Price Ladder). In each period t, the price p_t must be chosen from a *price ladder*, i.e., a set of admissible prices $\{q^0 > q^1 > \cdots > q^K\}$, where q^0 is the regular price and $q^1, \ldots, q^K$ are possible promotional prices.

We can model the business rule mathematically by writing the price at time t as follows:

$$p_t = \sum_{k=0}^{K} q^k \gamma_t^k, \tag{2}$$

where γ_t^k is a binary variable that is equal to 1 if the price q^k is selected at time t, and 0 otherwise. Therefore, instead of using the prices p_t as the set of decision variables, we use the set of binary variables $\{\gamma_t^k\colon t = 1, \ldots, T,\ k = 0, \ldots, K\}$, which is a total of $(K+1)T$ variables. To ensure that a single price is selected for each time period t, we impose the additional constraints:

$$\sum_{k=0}^{K} \gamma_t^k = 1 \quad \forall\, t.$$

Business Rule 1 is in contrast to the assumption made by other papers such as Popescu and Wu (2007), Kopalle et al. (1996), where the retailer can choose continuous prices. Note that, in practice, the retailer can only charge discrete prices. In the supermarket applications that we were involved in, this was an important business rule.[1] More precisely, the price for each item at each time period is selected from a discrete set of prices that consists of a regular price and levels of discounts. In supermarket applications, for example, these discounted prices have to end by 9 cents or sometimes by 5 cents.

Remark 2 (Extension to Time-Dependent Price Ladder). For clarity, in this paper, we make the simplifying assumption that the price ladder is time independent, i.e., one can charge $p_t = q^k$ for all t and k.

One can extend the analysis and results of this paper to the case where the price ladder is time dependent, i.e., the price ladder for period t is given by $\mathbb{Q}^t = \{q_t^0 > q_t^1 > \cdots > q_t^{K_t}\}$. Note that the number of permissible prices and the minimal price $q_t^{K_t}$ can be time dependent. In this case, Equation (2) becomes: $p_t = \sum_{k=0}^{K} q_t^k \gamma_t^k$.

Business Rule 2 (Limited Number of Promotions). The retailer may have to limit the number of promotions for a given product. This requirement is motivated from the fact that retailers wish to preserve the image of their store and not to train customers to be deal seekers. For example, it may be required to promote a particular product at most $L = 3$ times during the quarter. This constraint can be expressed mathematically as follows:

$$\sum_{t=1}^{T} \sum_{k=1}^{K} \gamma_t^k \leqslant L. \tag{3}$$

Business Rule 3 (Separating Periods between Consecutive Promotions). A common additional requirement is to space out two successive 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. In addition, this type of business requirement is often dictated directly by the manufacturer that wants to restrict the frequency of promotions to preserve the image of the brand. Mathematically, we have

$$\sum_{\tau=t}^{t+S}\sum_{k=1}^{K}\gamma_{\tau}^{k}\leqslant 1 \quad \forall\, t. \tag{4}$$

It is important to understand why, in practice, Business Rules 2 and 3 are commonly adopted by FMCG retailers to limit the frequency of price promotions. From an optimization point of view, these business rules appear to be unnecessarily restrictive. Mathematically, the larger the promotion limit L and the smaller the separating periods S, the larger the feasible region of the optimization problem, and therefore the greater the optimal profit. Although relaxing or removing Business Rules 2 and 3 will increase short-term profitability, retailers still follow these business rules as they recognize that running promotions too frequently, can hurt their long-term profit. Frequent promotions can negatively affect a retailer's brand image by conditioning consumers to perceive regular prices as bad deals. Finally, we note that, in some cases, Business Rules 2 and 3 may be soft constraints. In this case, one can perform a sensitivity analysis by solving the POP for different values of L and S (see such an example in Section 7.3). If a slight change in the parameters could lead to a significant increase in profit, upper management could be convinced to relax the constraints on L and S by renegotiating with the appropriate manufacturer.

3.3. Problem Formulation

We next present our formulation of the single-item POProblem:

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

where:

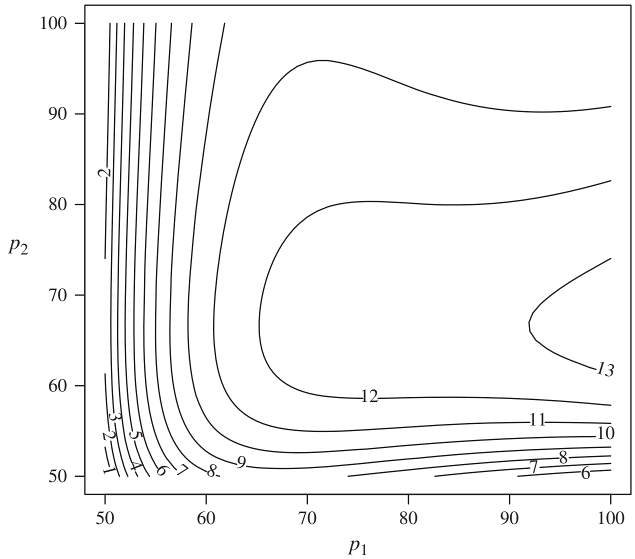
- T—Number of weeks in the horizon (e.g., one quarter composed of 13 weeks).
- L—Limitation on the number of promotions.
- S—Number of separating periods (separation time between two successive promotions).
- $\mathcal{Q}=\{q^0>q^1>\cdots>q^k>\cdots>q^K\}$—Price ladder, i.e., the discrete set of admissible prices.
- q^0—Regular (nonpromoted) 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.

Note that the only decisions are which price to choose from the price ladder at each time (i.e., the binary variables γ_t^k). We denote by POP($\mathbf{p}$) (or equivalently, POP(γ)) 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 as well as on demand seasonality and trend. Specific examples are presented in Section 5.

Remark 3 (End-of-Horizon Effects). The POP formulation above may be affected by the *end-of-horizon effect* (see, e.g., Herer and Tzur 2001). More specifically, the post-promotion dip in demand induces the promotions in periods $t\in(T-M,T]$ to reduce the demand in periods $t\in(T,T+M]$. Since the POP only considers the demand during periods $t\in[1,T]$, it ignores the demand reduction caused by promotions in periods $t\in(T-M,T]$. As a result, it creates an artificial advantage to schedule promotions at the end of the horizon. One of the methods to eliminate the end-of-horizon effect is to modify the formulation (POP) by extending the time horizon from $[1,T]$ to $[1,T+M]$, and adding the constraints $p_t=q^0$ for $t\in(T,T+M]$. The modified formulation takes into account the full effect of a post-promotion dip in demand for promotions in periods $t\in(T-M,T]$, thus eliminating the end-of-horizon effect (a similar argument applies for the beginning horizon effect). For simplicity, we focus in the remainder of the paper on the POP formulation above, rather than the modified version just described. We note that the analysis and results remain valid for the modified formulation.

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 continuous values, the objective is, in general, neither concave nor convex due to the cross-time dependence between prices (see Figure 3). Second, even if the objective was 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 a demand with post-promotion dip effect



Notes. Parameters: The demand functions at time 1 and 2 are given by $d_1(p_1, 100) = 1{,}000 \cdot p_1^{-4} \cdot 100^2$, $d_2(p_2, p_1) = 1{,}000 \cdot p_2^{-4} \cdot p_1^2$. The maximum and minimum prices are $q^0 = 100$ and $q^K = 50$, respectively, and the costs are $c_1 = c_2 = 50$.

4. IP Approximation

By looking carefully at several data sets, we observed that for many products, the promotions often last for one week, and that two consecutive promotions are at least three weeks apart. If the promotions are subject to a separating constraint as in (4), then the interaction between successive promotions should be fairly weak. Therefore, by ignoring the second-order interactions between promotions and capture only the direct effect of each promotion separately, we introduce a linear IP formulation that yields a "good" solution. More specifically, we approximate the nonlinear POP objective by a linear approximation based on the sum of unilateral deviations.

To derive the IP formulation of the POP, we first introduce some additional notation. For a given price vector $\mathbf{p} = (p_1, \ldots, p_T)$, we define the function (which is also named POP in a slight abuse of notation) that computes the total profit throughout the horizon:

$$\mathrm{POP}(\mathbf{p}) = \sum_{t=1}^{T} (p_t - c_t) d_t(\mathbf{p}_t). \tag{5}$$

We define the price vector $\mathbf{p}_t^k$ (with T elements) as follows:

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

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

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

These coefficients represent the unilateral deviations in total profit by applying a single promotion. One can compute each of these TK coefficients before starting the optimization procedure. Since one can do these calculations done offline, it does not affect the optimization complexity. We are now ready to formulate the IP approximation of the POP:

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

We make the following observations about the (IP) problem.

Observation 1. The constraints in (IP) are identical to the constraints of the original problem (POP). Consequently, the two problems have the same feasible region.

Observation 2. The business rules from the constraint set are modeled as linear constraints. Consequently, the IP formulation is a linear problem with integer variables.

Observation 3. The objective function in (IP) is a linear approximation of the objective in (POP). More precisely, it is a first-order discrete Taylor expansion of the the (POP) objective around the point $\gamma_t^0 = 1$ for $t = 1, 2, \ldots, T$.

In a slight abuse of notation, we define a function also named IP:

$$\mathrm{IP}(\mathbf{p}) = \sum_{t=1}^{T} \sum_{k=1}^{K} b_t^k \gamma_t^k, \tag{7}$$

where γ_t^k is such that $p_t = \sum_{k=1}^{K} q^k \cdot \gamma_t^k$. Note that although the IP function is linear in the binary variable space γ_t^k, in general it is not linear in the price space $\mathbf{p}$.

Observation 4. The objective function in (IP) captures the postpromotion dip effect but neglects the effect of interactions between two or more promotions. In other words, the POP and IP functions coincide at price vectors with zero or a single promotion, but may diverge at price vectors with two or more promotions.

Observation 5. We next present a reformulation of (IP) based on the following observation. Let us define the binary decision variables $\gamma_t = \sum_{k=1}^{K} \gamma_t^k$, and the coefficients $\tilde{b}_t = \max_{k=1,\ldots,K} b_t^k \ \forall\, t = 1, \ldots, T$. Then, (IP) is equivalent to the following compact formulation:

$$\begin{aligned} \max_{\gamma_t^k} \quad & \sum_{t=1}^{T} \tilde{b}_t \gamma_t, \\ \text{s.t.} \quad & \sum_{t=1}^{T} \gamma_t \leqslant L, \\ & \sum_{\tau=t}^{t+S} \gamma_\tau \leqslant 1, \quad \forall\, t, \\ & \gamma_t \in \{0,1\}, \quad \forall\, t. \end{aligned} \tag{CIP}$$

The compact integer programming (CIP) formulation (CIP) provides the following insight. One can separate the two types of decisions: (i) the promotion timing and (ii) the promotion depth. In particular, one can preprocess and decide the best promotion depth at each time period (i.e., if we end up deciding to promote at time t, we will use the promotion price p_t^k) by picking the highest value of b_t^k over $k = 0, 1, \ldots, K$ at each period t. Then, one can optimally decide the promotion scheduling by solving problem (CIP).

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 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* ($S \geqslant 1$)*, and the 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.,* $S \geqslant M$).

Proof of Proposition 1. (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 deviations.

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

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

Note that, in this case, the promotions affect only the current and next period demands, but not the demand in periods $t+2, t+3, \ldots, T$. 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 next show that the profit from having both promotions is equal to the sum of the incremental profits from each promotion separately; that is,

$$\begin{aligned} &\mathrm{POP}(\mathbf{p}\{p_t = q^i, p_u = q^j\}) - \mathrm{POP}(\mathbf{p}^0) \\ &\quad = \mathrm{POP}(\mathbf{p}\{p_t = q^i\}) - \mathrm{POP}(\mathbf{p}^0) \\ &\qquad + \mathrm{POP}(\mathbf{p}\{p_u = q^j\}) - \mathrm{POP}(\mathbf{p}^0). \end{aligned} \tag{8}$$

(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 $S \geqslant M$, the IP approximation is not neglecting correlations between different promotions, and hence optimal. □

In general, solving an IP can be difficult from a computational complexity standpoint. In our numerical experiments, we observed that Gurobi solves (IP) (or problem (CIP)) in less than a second. This follows from the fact that (IP) solves quickly has a feasible region that is an integral polyhedron.

Proposition 2. *The optimization problem* (CIP) *admits an integral feasible region.*

Proof of Proposition 2. Consider the problem (CIP). Observe that the constraint matrix of the feasible region has the consecutive ones property, and hence is totally unimodular. As a result, the formulation is integral. □

Given that (CIP) has a feasible region that is an integral, we can solve (CIP) efficiently by solving the LP relaxation:

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

5. Demand Models

This section is organized as follows. In Section 5.1, we empirically motivate the post-promotion dip effect by using supermarket data to estimate the commonly used log–log demand model. We then propose three classes of demand functions that capture the post-promotion dip effect, and admit as special cases several models used, in practice. In Section 5.2, we study a general class of demand functions where past prices have a multiplicative effect on current demand. In Appendix EC.5, we consider demand functions where past prices have additive effects on current demand. For each class of demand functions (multiplicative and additive), we exploit the structure of the function to derive bounds on the quality of the LP approximation.

5.1. Empirical Motivation

In this section, we empirically study the post-promotion dip effect by analyzing supermarket data. The analysis presented here is brief—a more detailed exposition on the data and estimation methods can be found in Section 7.

Our data set consists of 117 weeks of sales data from a supermarket for several brands of ground coffee. We divide our data in a training set of 82 weeks and a testing set of 35 weeks. Our basic demand model is the log–log demand model in (9), which is commonly used in industry (for example, by Oracle Retail) and in academia (see, e.g., Heerde et al. 2000, Macé and Neslin 2004). We estimate two versions of the model with different values of the memory parameter M.

- Model 1 is estimated with $M = 0$ in (9). This corresponds to a model without the post-promotion dip effect, so the current demand d_t depends only on the current price p_t and not on past prices.
- Model 2 is estimated with $M = 2$ in (9). This corresponds to a model with a post-promotion dip effect, i.e., the current demand d_t depends on the current price p_t and on the prices in the two prior weeks p_{t-1} and p_{t-2}.

After estimating the demand model parameters using the training set, we predict the sales for the test set and compute the forecast accuracy metrics. The forecast metrics for the product labeled as "Brand1"[2] are shown in Table 1. The forecast metrics (MAPE, OOS R^2, and revenue bias) are formally defined in Section 7. It can be seen from Table 1 that the forecast accuracy of Model 2 is significantly higher relative to the forecast accuracy of Model 1. In addition, the regression statistics for Model 2 are shown in Table 2. We note that the predicted sales for "Brand1" is given by the equation:

$$\log d_t = \beta^0 + \beta^1 t + \beta^2 \text{WEEK}_t - 3.277 \log p_t + 0.518 \log p_{t-1} + 0.465 \log p_{t-2}, \tag{9}$$

where β^0 and β^1 denote the brand intercept and the trend coefficient, respectively, and $\beta^2 = [\beta_t^2]; t = 1, \ldots, 52$ is a vector of weekly seasonality coefficients. Table 2 shows that the price elasticity coefficients of p_{t-1} and p_{t-2} for Model 2 are statistically significant. Finally, since the coefficients of the past prices p_{t-1} and p_{t-2} are positive, we conclude that the post-promotion dip property holds for the demand function in (9).

Table 1. Forecast accuracy for "Brand1" (ground coffee)

Variable	Model 1 ($M = 0$)	Model 2 ($M = 0$)
MAPE	0.145	0.116
OOS R^2	0.827	0.900
Revenue bias	1.069	1.059

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

Variable	Coefficient	Std. error	p-value
*Brand*1			
$\log p_t$	−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*
*Brand*2			
$\log p_t$	−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
*Brand*1			
MAPE		0.116	
OOS R^2		0.900	
Revenue bias		1.059	
*Brand*2			
MAPE		0.097	
OOS R^2		0.903	
Revenue bias		1.017	

Notes. IS adjusted R^2. Significance codes: *indicates significance <0.05, ***indicates significance < 0.001.

5.2. Multiplicative Demand

In the following, we refer to (IP) as the LP approximation and denote its optimal solution by γ^{LP}. In addition, LP($\mathbf{p}$) (or, equivalently, LP(γ)) denotes the objective function of (LP) evaluated at the vector $\mathbf{p}$ (or equivalently γ).

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 as

$$d_t = f_t(p_t) \cdot g_1(p_{t-1}) \cdot g_2(p_{t-2}) \cdot \ldots \cdot g_M(p_{t-M}). \tag{10}$$

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. The parameter M represents the memory of consumers with respect to past prices and can be estimated from data. As we verify in Section 7 using actual data, it is reasonable to assume the following for the functions $g_k(\cdot)$.

Assumption 4 (Conditions for Multiplicative Demand).
1. *Past promotions have a multiplicative reduction effect on current demand, i.e.,* $0 < g_k(p) \leqslant 1$.
2. *Deeper promotions result in larger reduction in future demand, i.e., for* $p \leqslant q$, *we have*: $g_k(p) \leqslant g_k(q) \leqslant g_k(q^0) = 1$.
3. *The reduction effect is nonincreasing with time after the promotion*: g_k *is nondecreasing with respect to* k, *i.e.,* $g_k(p) \leqslant g_{k+1}(p)$.

In addition, we adopt the convention that $g_k(p) = 1$ for all $k > M$, so that no effects are present after M periods. We next discuss Assumption 4 in more detail. The nominal part of the demand $f_t(p_t)$ is assumed to be nonnegative so when the factors that depend on past prices are absent, the demand is nonnegative. The first requirement $g_k(p) \leqslant 1$ follows from the fact that promotions in past periods may reduce current demand (capturing the postpromotion dip effect). For example, the consumers can be reference dependent by looking at the difference between the current price p_t and the effects of the past M prices. The second part of Assumption 4 relates to the comparison effect of consumers. In particular, by comparing the current price p_t to the fact that prices were lower in the past, it creates a feeling of loss that reduces the current demand. In addition, this feeling of loss is larger when the past promotion is deeper. Finally, the third part may suggest that the more recent promotions have a higher impact on current demand relative to older promotions. This implies that the consumers' reference points are modeled in a similar fashion as an exponential smoothing.

Remark 4 (General Demand Model). The demand in (10) represents a general class of demand models, which admits as special cases several models used, in practice. For example, the demand model in Heerde et al. (2000) or a special case of the model in Macé and Neslin (2004) is of the general 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 (10).

5.2.1. Bounds on Quality of Approximation. We bound the difference in profit between the POP and LP solutions based on the effective maximal number of promotions, denoted by $\tilde{L}$:

$$\tilde{L} = \min\{L, \tilde{N}\}, \quad \text{where } \tilde{N} = \left\lfloor \frac{T-1}{S+1} \right\rfloor + 1. \tag{11}$$

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

Theorem 1. *Let* γ^{POP} *be an optimal solution to* (POP) *and let* γ^{LP} *be an optimal solution to* (LP). *Then,*

$$1 \leqslant \frac{\mathrm{POP}(\gamma^{\mathrm{POP}})}{\mathrm{POP}(\gamma^{\mathrm{LP}})} \leqslant \frac{1}{\underline{R}}, \tag{12}$$

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

$$\underline{R} = \prod_{i=1}^{\tilde{L}-1} g_{i(S+1)}(q^K), \tag{13}$$

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:

$$\begin{aligned}\underline{R} \cdot \mathrm{LP}(\gamma^{\mathrm{LP}}) &\overset{\text{(i)}}{\leqslant} \mathrm{POP}(\gamma^{\mathrm{LP}}) \overset{\text{(ii)}}{\leqslant} \mathrm{POP}(\gamma^{\mathrm{POP}}) \\ &\overset{\text{(iii)}}{\leqslant} \mathrm{LP}(\gamma^{\mathrm{POP}}) \overset{\text{(iv)}}{\leqslant} \mathrm{LP}(\gamma^{\mathrm{LP}}).\end{aligned}$$

Inequality (i) follows from Proposition 3. Inequality (ii) follows from the optimality of γ^{POP}, and inequality (iii) follows from part 2 of Lemma 1. Finally, inequality (iv) follows from the optimality of γ^{LP}. Therefore we obtain:

$$\begin{aligned}\underline{R} &= \underline{R} \cdot \frac{\mathrm{POP}(\gamma^{\mathrm{POP}})}{\mathrm{POP}(\gamma^{\mathrm{POP}})} \leqslant \underline{R} \cdot \frac{\mathrm{LP}(\gamma^{\mathrm{LP}})}{\mathrm{POP}(\gamma^{\mathrm{POP}})} \\ &\leqslant \frac{\mathrm{POP}(\gamma^{\mathrm{LP}})}{\mathrm{POP}(\gamma^{\mathrm{POP}})} \leqslant \frac{\mathrm{POP}(\gamma^{\mathrm{POP}})}{\mathrm{POP}(\gamma^{\mathrm{POP}})} \\ &= 1. \quad \square\end{aligned}$$

Theorem 1 relies on Lemma 1 and Proposition 3. Before stating Lemma 1, we first introduce the following notation.

Let $\mathcal{A} = \{(t_1, k_1), \ldots, (t_N, k_N)\}$ with $N \leqslant L$ be a set of promotions with $1 \leqslant t_1 < t_2 < \cdots < t_N \leqslant T$. In other words, at each time period t_n; $\forall\, n = 1, \ldots, N$ the promotion price q^{k_n} is used, whereas at the remaining time periods, the regular price q^0 (no promotion) is set. We define the price vector associated with the set $\mathcal{A}$ as

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

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

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

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

Note that matrix $(\gamma_{\mathcal{A}})_t^k$ has dimensions $(K+1)\times T$. In the previous example, we have

$$\gamma_{\mathcal{A}} = \begin{bmatrix} 0 & 1 & 0 & 1 & 1 \\ 1 & 0 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 \end{bmatrix},$$

Recall that the LP objective function is given by

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

where b_t^k is defined in (6).

Lemma 1 (Submodular Effect of the Last Promotion on Profits). 1. *Let $\mathcal{A} = \{(t_1, k_1), \ldots, (t_n, k_n)\}$ be a set of promotions with $t_1 < t_2 < \cdots < t_n$ $(n \leqslant L)$ and let $\mathcal{B} \subset \mathcal{A}$. Consider a new promotion (t', k') with $t_n < t'$. If the new promotion (t', k'), when added to $\mathcal{A}$, yields a larger profit than $\mathbf{p}_{\mathcal{A}}$; that is,*

$$\mathrm{POP}(\gamma_{\mathcal{A}\cup\{(t',k')\}}) \geqslant \mathrm{POP}(\gamma_{\mathcal{A}}), \tag{14}$$

then the promotion (t', k') yields a larger marginal profit increase for $\mathbf{p}_{\mathcal{B}}$ than for $\mathbf{p}_{\mathcal{A}}$; that is,

$$\begin{aligned}&\mathrm{POP}(\gamma_{\mathcal{A}\cup\{(t',k')\}}) - \mathrm{POP}(\gamma_{\mathcal{A}}) \\ &\quad\leqslant \mathrm{POP}(\gamma_{\mathcal{B}\cup\{(t',k')\}}) - \mathrm{POP}(\gamma_{\mathcal{B}}).\end{aligned} \tag{15}$$

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

Note that Lemma 1 does not guarantee that the subadditivity property (15) holds for a general feasible solution γ, but only if γ satisfies condition (14). Fortunately, the required condition in (14) is always automatically satisfied for the optimal POP solution. The proof of Lemma 1 can be found in Appendix EC.1. Lemma 1 states that for a multiplicative demand model as in (10), the POP profit is 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., $\mathrm{POP}(\gamma^{\mathrm{POP}}) \leqslant \mathrm{LP}(\gamma^{\mathrm{POP}})$.

By using Lemma 1, one can see that the POP profit is submodular in the number of promotions. This means that the marginal effect on an additional promotion has a smaller impact when we already have many promotions in the selling season. As a result, scheduling correctly the first few promotions is very important for the retailer and a myopic solution (that does not take into account the future time periods) may perform badly.

The main insight from this result is as follows. Recall that we have shown the submodularity property for multiplicative demand models, whereas for additive demand, the POP profits are supermodular in promotions (as we will show in Appendix EC.5). One can naturally believe (and most of category managers we interacted with share this intuition) that the true effect observed, in practice, is closer to submodular. This follows from the fact that, if there are many promotions, the effect on profit is not as large as for the first few promotions, where many consumers will switch and buy the product. Consequently, to capture this feature, one should consider a multiplicative demand model. In particular, the linear demand model (that is commonly used in many applications) is not appropriate in this setting. Interestingly, most of the additive demand models we tried did not fit the data as well as multiplicative demand models. One possible explanation (that the additive models do not yield a good fit to the data) may come from the fact that the POP profit is supermodular, whereas, in practice, it is submodular (and hence, a multiplicative model is more suitable).

Proposition 3. *For any feasible vector γ, we have* $\mathrm{POP}(\gamma) \geqslant \underline{R} \cdot \mathrm{LP}(\gamma)$.

The proof of Proposition 3 can be found in Appendix EC.2. It provides a lower bound for the POP objective by applying the linearization and compensating by the worst-case aggregate factor $\underline{R}$.

Using Theorem 1, one can solve the LP approximation (efficiently) and obtain a guarantee relative to the optimal POP solution. These bounds are parametric and can be applied to any general demand model in the form of (10). In addition, as we illustrate in Section 5.2.2, these bounds perform well, in practice, for a wide range of parameters. We next show that the bounds of Theorem 1 are tight.

Proposition 4 (Tightness of the Bounds for Multiplicative Demand). 1. *The lower bound in Theorem 1 is tight. More precisely, for any given price ladder, L, S, and functions g_k, there exist T, costs c_t, and functions f_t such that*

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

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

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

where we denote the bound with n promotions by: $\underline{R}_n = \prod_{i=1}^{n-1} g_{i(S+1)}(q^K)$, with $\underline{R}_0 = 1$ by convention. We then define the following limit: $\underline{R}_\infty = \lim_{n\to\infty} \underline{R}_n$.

The proof of Proposition 4 can be found in Appendix EC.3.

Note that our analytical bound $\underline{R}$ from Equation EC.3 depends only on the minimal element of the price ladder q^K. As a result, it provides a guarantee that does not depend on the (unknown) optimal pricing policy. One advantage is that one can evaluate the bound very easily. The drawback is that this guarantee can potentially be far from the attained profit (even though the bound is tight, as we show in Proposition 4). If we happen to have further knowledge about the optimal solution, one can then improve the bound. For example, if we know that the minimal price will be used at most twice, one can incorporate this information and obtain a better refined bound. However, in most cases, it is very hard to have some trustful information about the optimal pricing policy. Consequently, our bound provides a performance guarantee that does not depend on the optimal policy. This observation also supports the fact that the actual performance, i.e., the actual profit ratio $\mathrm{POP}(\gamma^{\mathrm{LP}})/\mathrm{POP}(\gamma^{\mathrm{POP}})$ is most of the time closer to 1 relative to $\underline{R}$, as we illustrate next.

5.2.2. Discussing the Bounds. We summarize the main findings regarding the behavior and quality of the bounds we have developed in the previous section. Recall that solving the POP can be hard, in practice, and one can instead implement the LP solution. The resulting profit is equal to $\mathrm{POP}(\gamma^{\mathrm{LP}})$, whereas in theory, we could have obtained a maximum profit of $\mathrm{POP}(\gamma^{\mathrm{POP}})$. In our computational experiments, we examine the gap between $\mathrm{POP}(\gamma^{\mathrm{LP}})$ and $\mathrm{POP}(\gamma^{\mathrm{POP}})$ as a function of the various problem parameters. In addition, we compare the ratio between $\mathrm{POP}(\gamma^{\mathrm{POP}})$ and $\mathrm{POP}(\gamma^{\mathrm{LP}})$ relative to the upper bound in Theorem 1 equal to $1/\underline{R}$. As we previously noted, the bounds depend on four different parameters: the number of separating periods S, the number of promotions allowed L, the effect of past prices (i.e., the value of the memory parameter M as well as the magnitude of the functions $g_k(\cdot)$), and the minimal price q^K. Below, we summarize the effect of each of these factors for the following demand: $\log d_t(\mathbf{p}) = \log(10) - 4\log p_t + 0.5\log p_{t-1} + 0.3\log p_{t-2} + 0.2\log p_{t-3} + 0.1\log p_{t-4}$, with $T = 9$.

The details of the tests are presented in Appendix EC.4 and are summarized here: (a) In most cases, the LP solution achieves a profit that is very close to optimal. In particular, the actual optimality gap (between the POP objective at optimality versus evaluated at the LP solution) seems to be of the order of 1%–2% and is much smaller than the upper bound in Theorem 1. (b) The upper bound $1/\underline{R}$ varies between 1 and 1.33 depending on the values of the parameters. (c) As S increases, the upper bound $1/\underline{R}$ improves. Indeed, the promotions are further apart in time, reducing the interaction between promotions, and improving the quality of the LP approximation. For values of $S \geqslant 1$, the upper bound is at most 1.11 in this example. In practice, typically the number of separating periods is at least 1 but often two to four weeks. (d) For values of L between 1 and 8, the upper bound is at most 1.23 in this example. (e) The upper bound decreases with q^K and is at most 1.32, when a 50% promotion is allowed. If we restrict to a maximum of 30% promotion price, the bound becomes 1.14. (f) The upper bound increases with the memory parameter M and is at most 1.23 in this example.

In each of our experiments, the profit of the LP solution is close to the profit of the optimal solution. The maximum observed theoretical bound on the profit ratio was below 1.35, whereas the maximum observed actual profit ratio was below 1.02. Equivalently, the theoretical bound predicts that the LP solution will attain 82% of the profit of the optimal POP solution in the worst case, whereas, in practice, the LP solution attains approximately 99% of the optimal profit. We observed that in many cases, the actual profit ratio was significantly better relative to the theoretical bound.

To explain why the LP solution provides such a good profit ratio, we next consider two different scenarios characterized by the strength of the post-promotion dip effect. Recall that the LP approximation can be viewed as a first-order Taylor expansion around the regular price. Therefore the LP objective captures exactly the effect of any single promotion, but neglects the higher-order terms, which capture the interaction of multiple promotions.

1. In the first scenario, the post-promotion dip effect is weak, i.e., the memory parameter M is small and/or the functions $g_k(\cdot)$ are close to 1. It is clear that for such cases, the LP approximation performs well, because the interaction between multiple promotions is weak.

2. In the second scenario, the post-promotion dip effect is strong, i.e., the memory parameter M is large and/or the functions $g_k(\cdot)$ are not close to 1. In this case, the terms we neglect (interactions between promotions) can be significant. However, due to a strong post-promotion dip effect, it becomes optimal to space out the promotions in time, as a promotion will reduce future demand. Consequently, the strong post-promotion dip effect drives the optimal solution to automatically space out the promotions. When the promotions are far from each other, our LP approximation performs well; this is because the further two promotions are apart from each other, the weaker their interactions through the functions $g_k(\cdot)$ (see Assumption 4).

In conclusion, the LP approximation performs well in both regimes due to the nature of the post-promotion dip effect that drives the optimal solution in a good direction (i.e., to space out promotions). In addition, this LP approximation gives rise to interesting insights, which are discussed in Section 7.

6. Extension to Uncertain Demand

In this section, we extend our solution approach for the case where the demand is uncertain. We next discuss the analysis and results.

We assume that the demand function at time t, $d_t(p_t, p_{t-1}, \ldots, p_{t-M})$ can be one of J different scenarios (e.g., different functional forms or a single function with different parameters values). We denote each scenario by $d_t^j(\cdot)$; $\forall\, j = 1, \ldots, J$. For example, one can fit several structural forms to the data (e.g., log–log and linear). Alternatively, one can consider a single form (such as log–log), estimate the model parameters, and then assume that each estimated parameter lies within some given confidence interval (see a concrete example in Section 7.4). Our goal is to solve the POP for this setting. We consider two different types of objectives: (i) a robust formulation that maximizes the worst-case scenario over the J instances and (ii) an expectation formulation, where we maximize the expected profit weighted by the probability of each demand scenario.

The robust formulation is given by

$$\mathrm{POP}^R = \max_{p_1, p_2, \ldots, p_T} \min_{j=1,\ldots,J} \mathrm{POP}_j(p_1, p_2, \ldots, p_T), \tag{16}$$

subject to the usual constraints. Here, $\mathrm{POP}_j(p_1, p_2, \ldots, p_T)$ corresponds to the total profit over the horizon T, where one uses the demand function from scenario j. Note that the constraint set is not affected by the scenario and is the same for all $j = 1, \ldots, J$.

We propose the following method to solve problem (16). First, solve the POP problem for each scenario $j = 1, \ldots, J$ separately to obtain J solutions, denoted by p_j^*. Since each scenario is solved by the LP approximation, this can be done efficiently. Second, evaluate each objective POP_k at each solution p_j^*. Consequently, we have J^2 such evaluations. Third, identify the minimal value of the J^2 evaluations. The solution p_ℓ^* that attains this minimal value is called the robust solution and the corresponding objective is denoted by POP^R. This way, one can obtain a solution that is robust for any of the J scenarios and will account for demand uncertainty. Note that this method of solving problem (16) is not always optimal, but rather an efficient heuristic (since the objective is not linear, one cannot directly use robust optimization techniques to solve a single LP). Note also that one can extend the analytical bound from Theorem 1 to this setting. In particular, one can compute the bound for each scenario $j = 1, \ldots, J$ separately, denoted by $\underline{R}_j$. Then, the minimal bound over all the J scenarios yields a bound for problem (16).

One concern with the above method is that it may be too conservative. We address this concern in Section 7.4, where we test this approach using sales data from a supermarket retailer.

The expectation formulation is given by

$$\mathrm{POP}^A = \max_{p_1, p_2, \ldots, p_T} \sum_{j=1}^{J} \mathrm{prob}_j\, \mathrm{POP}_j(p_1, p_2, \ldots, p_T), \tag{17}$$

subject to the usual constraints. Here, prob_j corresponds to the probability (or relative confidence) of scenario j ($\sum_{j=1}^{J} \mathrm{prob}_j = 1$). For example, the probability prob_j can be computed by using the relative values of some forecast metrics such as R^2 and MAPE for each scenario. Note that when $\mathrm{prob}_j = 1$ for a particular scenario, problem (17) reduces to the basic problem without any demand uncertainty. Note also that one can show that $\mathrm{POP}^A \geqslant \mathrm{POP}^R$ for any given probability distribution vector prob_j; $\forall\, j = 1, \ldots, J$. Finally, one can see that problem (17) is equivalent to the POP when the demand function is replaced by the expected demand given by $\sum_{j=1}^{J} \mathrm{prob}_j d_t^J(\cdot)$. Consequently, one can apply the LP approximation method to solve problem (17) and derive the analytical bound on the quality of the approximation.

In conclusion, we proposed two methods (that require to solve a small number of LPs) that allow us to handle demand uncertainty for our problem. We will test and compare the two methods using our real-world example in Section 7.4 and demonstrate that our solution is robust to demand forecast errors.

7. Case Study

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, estimate the demand model we introduce, validate it, compute the optimized prices from our LP approximation, 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 (10) 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 time stamp, and the purchased items during that transaction. 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 the 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 (9)):

$$\log d_{it} = \beta^0\,\mathrm{BRAND}_i + \beta^1 t + \beta^2\,\mathrm{WEEK}_t + \sum_{m=0}^{M} \beta^3_{im} \log p_{i,t-m} + \epsilon_t, \qquad (18)$$

where i and t denote the brand and time indices, d_{it} denotes the sales (which are equal to the demand, as we discussed in Section 3) of brand i in week t, BRAND_i and WEEK_t denote brand and week indicators, and p_{it} corresponds to the average per unit selling price of brand i in week t. β_0 and β_2 are vectors with components for each brand and each week, respectively, whereas β_1 is a scalar that captures the trend. Note that the seasonality parameters β_2 for each week of the year are jointly estimated across all the brands in the category. The additive noises ϵ_t; $\forall\, t = 1, \ldots, T$ account for the unobserved discrepancies and are assumed to be normally distributed and i.i.d. Similar demand models have been used in the literature, e.g., Heerde et al. (2000), Macé and Neslin (2004).

The demand in (18) 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 $M = 0$, only the current price affects the demand in week t. When the memory parameter $M = 2$, then the demand in week t depends on the current price p_t, and also on the price of the two previous weeks p_{t-1} and p_{t-2}. The memory parameter M is estimated from data as follows. We first incorporate all the past prices (i.e., $p_{t-1}, p_{t-2}, \ldots, p_{t-T}$) in the regression model and observe that only the M first ones (depending on the type of product and the specific store, M was between 1 and 4) were statistically significant, i.e., the p-value was less than 0.05 (see Table 2). We then remove the nonsignificant observable variables and re-estimate the model parameters. We note that our model does not account explicitly for cross-brand effects, i.e., we assume that the demand for brand i depends only on the prices of brand i. This assumption is reasonable for certain products such as coffee because people are loyal about the brand they consume, and do not easily switch between brands. In addition, the high predictive accuracy of our model validates this assumption.

For ease of notation, from this point, we drop the brand index i since we estimate and optimize for a single-item model. Observe that one can define

$$f_t(p_t) = \exp\big(\beta^0 + \beta^1 t + \beta^2\,\mathrm{WEEK}_t + \beta^3_0 \log p_t\big),$$
$$g_m(p_{t-m}) = (p_{t-m})^{\beta^3_m}, \quad m = 1, \ldots, M,$$

and therefore Equation (18) is a special case of the multiplicative model in (10).

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., $\beta^3_0 < 0$.
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., $\beta^3_m \geqslant 0$ 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., $\beta^3_m > \beta^3_{m+1}$ for $m = 1, \ldots, M-1$.

We note that the conditions above are a special case of Assumption 4 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 sales using the test set. To measure the forecast accuracy, we use the following forecast metrics. In the sequel, we use the notation d_t for the actual sales (or equivalently demand) and $\hat{d}_t$ be the forecasted values.

- The MAPE is given by

$$\mathrm{MAPE} = \frac{1}{T}\sum_{t=1}^{T} \frac{|d_t - \hat{d}_t|}{d_t}.$$

The MAPE captures the average relative forecast error in absolute value.

- The R^2 is given by

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

where $\bar{d} = \sum_{t=1}^{T} d_t/T$, $SS_{\mathrm{tot}} = \sum_{t=1}^{T}(d_t - \bar{d})^2$, and $SS_{\mathrm{res}} = \sum_{t=1}^{T}(d_t - \hat{d}_t)^2$. We distinguish between in-sample (IS) and OOS R^2. In addition, we consider the adjusted R^2 to account for the number of explanatory variables

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

where p is the total number of independent variables (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

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

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., Gupta 1988 and Villas-Boas 1995). We use a linear regression to estimate the parameters of the demand model in (18) for five different coffee brands. 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 for the test set in Figure 4. Remember that our data consists of 117 weeks, which we split into 82 weeks on training and 35 weeks of testing.

On one hand, Brand1 is a private label coffee brand, which has frequent promotions (approximately once every four weeks). The price elasticity coefficients for the current price and two previous prices are statistically significant suggesting that for this brand, the memory parameter is $M = 2$. In contrast, Brand2 is a premium coffee brand, which has also frequent promotions (approximately once every five weeks). The price elasticity coefficients for 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 is $M = 1$. By looking at the statistically significant price coefficients, one can see 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 in the chocolate, tea, and yogurt categories. The results are summarized in Table 3. We do not report the individual product coefficients but we note that they follow our expectations in terms of sign and ordering. We highlight that the forecast error is low as evidenced by the high IS and OOS R^2, the low MAPE values, and a revenue bias being close to 1.

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 goods (such as yogurt), consumers do not stockpile, and therefore the memory parameter is zero. However, coffee is clearly a less perishable product, and hence stockpiling is more significant.

Figure 4. Actual vs. forecasted sales over the 35 test weeks for Brand1

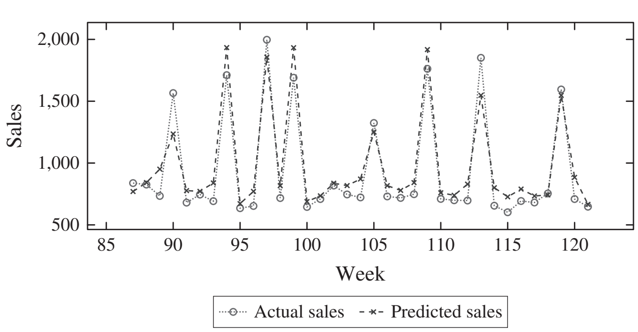


Table 3. Summary of the estimation results for four product categories

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

Note. The column for product memories indicates all values of the memory parameter that we found for the products in that category.

7.3. Optimization Results and Discussion

Having validated the forecasting demand model, we next compute and test the optimized promotion prices. We assume that the demand forecast is the true demand model, and use it as an input to our promotion optimization formulation (POP).

Experimental setup. We compute the optimal LP prices for Brand1 over the test set, i.e., $T = 35$ 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 one week). As stated earlier, the regular price is normalized to be 1. 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 that the cost of the product is constant, $c_t = 0.4$. Since the lowest price charged by the retailer was 0.75, the set of permissible normalized prices is chosen to be $\{0.75, 0.80, 0.85, \ldots, 1\}$. The LP optimization results are shown in Figure 5. We make the following observations.

- The predicted profit using the 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.

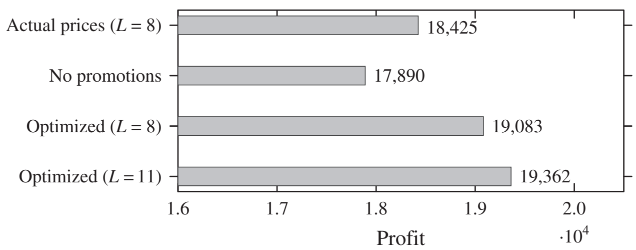


Figure 5. Profits for different scenarios using a log–log demand

• The predicted profit using the optimized LP prices imposing the same number of promotions as a business requirement ($L = 8$ during the 35 weeks) is \$19,083. This is a 3.5% gain relative to the benchmark. Therefore the estimated log–log model predicts that the optimal LP prices with the same number of promotions yield a 3.5% gain relative to the actual implemented profit. In other words, by only carefully planning the same number of promotions, our model suggests that the retailer can increase its profit by 3.5% in this case.

• The predicted profit using the optimal 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.

We next compute the bound from Theorem 1 for the actual data we have been using above. The lower bound can be rewritten as $\underline{R} \cdot \mathrm{POP}(\gamma^{\mathrm{POP}}) \leqslant \mathrm{POP}(\gamma^{\mathrm{LP}})$, where $\underline{R} = \prod_{i=1}^{\tilde{L}-1} g_{i(S+1)}(q^K)$, and therefore depends on the problem parameters. 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 $S \geqslant 2$, 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-possible case. 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 nonmonotone submodular function subject to no constraints admits a $1/2$ approximation algorithm (see, for example, Buchbinder et al. 2012, 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 parameters values. Recall also that, in practice, the LP approximation usually outperforms the bounds.

Next, we compare the running time of the LP to a naïve approach of using an exhaustive search method to find the optimal POP prices. Note that the POP objective is neither convex nor concave. The experiments were run using a desktop computer with an Intel Core i5-680 processor @ 3.60 GHz 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 two 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 three elements in the price ladder and $L = 8$, it requires three hours to solve, whereas the LP solution solves within milliseconds. We note that since we are considering nonlinear demand functions with integer variables, general methods to solve this problem do not exist in commercial solvers.

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 optimal prices, and one can compute the upper and lower bounds as a guarantee.

Finally, we use our model to infer the following useful rules for promotions:

• The less perishable the product is, the larger the value of M is and the stronger the post-promotion dip effect. Consequently, our model will plan the promotions to be more spaced out. This suggests that retailers want to space out promotions for perishable products. The same conclusion applies as the brand of the product is less premium.

• In the data we have, we observed that seasonality effects are not very significant. As a result, the actual time of a promotion (for this type of products) is not as important as the time relative to the previous or the next promotion for the same product.

• Carefully planning promotions can have an important impact on the retailer's profit and on the bottom line. In addition, the solution approach we proposed in this paper is robust to the demand parameters, as we have shown that the profit gain dominates the demand forecasting error in our case study. Therefore, retailers should carefully decide promotions and preferably with the help of decision support tools.

• Demand functions with multiplicative (additive) past prices effects induce the total profit to be submodular (supermodular) with respect to the number of promotions. When a retailer considers several functional forms to forecast demand, knowing that a multiplicative function (such as the log–log demand) yields a better fit relative to additive, this informs the retailer to be more cautious about planning carefully the first few promotions of the selling season.

7.4. Uncertain Demand

In Section 6, we proposed two different approaches to address the case where the demand is uncertain and can be one of J different functions. The first method is based on a robust objective denoted by POP^R, whereas the second considers an expectation formulation with objective POP^A. Recall that $\text{POP}^A \geqslant \text{POP}^R$. We are interested in comparing the two objectives relative to the nominal case where the demand is deterministic (i.e., a single scenario). To this end, we perform the following computational experiments. We begin with our nominal log–log demand model estimated from real data (see Table 2). More precisely, for each demand parameter, we have access to the expected (nominal) value, as well as its standard deviation found from the regression. The most important parameters in our context are the coefficients that multiply $\log p_t$, $\log p_{t-1}$ and $\log p_{t-2}$. The other estimated parameters (such as the seasonality, the trend factor, and the intercept) are either more certain (due to the fact that they are estimated across multiple products/stores) or not statistically significant. Therefore, their exact values do not affect the optimal pricing decisions significantly. We then assume that the price sensitivity estimated parameters (i.e., the coefficients of $\log p_t$, $\log p_{t-1}$, and $\log p_{t-2}$) lie in some confidence interval $[-A, A]$ around their respective nominal values. The confidence interval is chosen according to the standard deviation of the regression results (see Table 2). Each scenario $j = 1, \ldots, J$ is generated by randomly picking the three price sensitivity parameters (coefficients that multiply $\log p_t$, $\log p_{t-1}$, and $\log p_{t-2}$) in their respective confidence intervals. Unless otherwise specified, we assume $\text{prob}_j = 1/J$; $\forall j$ (i.e., a uniform distribution). We then perform the following four tests:

1. Test 1. We set $A = \sigma$ for each of the three parameters (here, σ denotes the standard deviation of the estimated parameters from the third column of Table 2), and generate $J = 10$ scenarios. In other words, we consider 10 randomly generated instances where the nominal estimated parameters can be increased or decreased by at most one standard deviation. Interestingly, solving the nominal problem, the robust version and the expectation formulation lead to the same optimal prices. In addition, the objective functions are all very close (within 1.9%).
2. Test 2. We repeat Test 1 for $J = 100$ scenarios. In this case, we obtain that the robust and nominal solutions are identical. The expectation formulation yields a solution that is very similar to the robust and nominal solutions, except that two promotion prices are swapped. Considering these three formulations does not significantly change the solution objective value (in our case, by less than 0.015%).
3. Test 3. We now consider $J = 10$ scenarios but $A = 2\sigma$, i.e., a larger uncertainty in the estimated parameters. Once more, the expected and robust formulations yield the same optimal solution as the nominal formulation with a deterministic demand.
4. Test 4. We generate $J = 100$ independent samples of perturbed demand realizations, using each one of the three distributions: uniform, Gaussian, and exponential. Note that we use truncated distributions between plus or minus one standard deviation to obtain meaningful comparisons.[3] Each sample corresponds to a demand model where the three estimated parameters are within one standard deviation of their nominal values. We then solve the robust version and the expectation formulation for each distribution. In each case, we find the optimal vector of prices, and evaluate the total profit on the nominal demand. The results are summarized in Table 4. Note that all the profits are very close to each other. In addition, the vector of prices are very similar in all the cases. Hence, this confirms that our solution approach is indeed robust with respect to the distribution of the parameters.

The above findings allow us to handle demand uncertainty in our problem in an efficient way. In particular, by perturbing the estimated demand model, we still obtain the same optimal prices as in the nominal deterministic case (or at least a solution that is very close). In addition, one can see that the optimal solution of the nominal problem (when demand is assumed to be a deterministic function) seems to be quite robust to variations in the estimated parameters. As a result, this further supports the deterministic demand assumption in our problem.

Table 4. Profit for randomly perturbed demands using three distributions

Variable	Nominal	Uniform	Gaussian	Exponential
Robust	19,083	19,086	19,080	19,086
Average	19,083	19,083	19,083	19,083

We next perform a series of tests to demonstrate that the profit gain dominates the forecasting error. We consider the log–log demand model estimated from data (see the demand parameters in Table 2). As before, we assume that the true demand model is not exactly known, so that the three estimated parameters can be within one standard deviation from their expected values. In other words, the nominal model assumes that the parameters are $\bar{\beta}$, but in reality, the parameters are $\tilde{\beta} \in [\bar{\beta} - \sigma, \bar{\beta} + \sigma]$.

In this case, we perturb the estimated parameters and generate $J = 10{,}000$ independent samples of perturbed demand realizations. Each sample corresponds to a demand model where the three parameters ($\tilde{\beta}$) are within one standard deviation of their nominal values. For each sample, we compute the ratio of the total profit evaluated at the optimized prices (using the nominal model) $\Pi(\mathbf{p}^{\mathrm{LP}})$, divided by the profit at the implemented prices $\Pi(\mathbf{p}^{I})$. Here, Π is the total profit over the 35 weeks, p^I is the vector of implemented prices, and p^{LP} is the vector of suggested prices using our approach with the (deterministic) nominal demand model. Note that we do not optimize and solve the LP for each sample. Instead, we solve the LP only for the nominal demand model. We then plot the histogram of the profit ratio (see Figure 6), compute the median and the 75th percentile. We obtained that the suggested promotions from our model allow a positive profit gain for all the samples. The profit gain is between 2.3% and 5.1%, with a median improvement of 3.66% (equal to the nominal one with deterministic demand). In addition, the 75th percentile is equal to 4.18%. The results are summarized in Table 5. This allows us to convey a convincing argument that our approach is robust with respect to the estimated demand parameters, and that the profit gain clearly dominates the forecasting error.

Figure 6. Histogram of profit ratios for 10,000 samples of perturbed demands

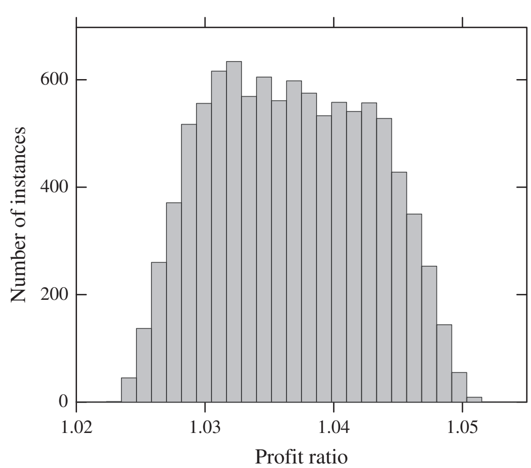


Table 5. Profit gain (in %) for randomly perturbed demands (using 10,000 independent samples)

Nominal	Minimum	25th percentile	Median	75th percentile	Maximum
3.66	2.3	3.17	3.66	4.18	5.1

We next perform a different test that confirms the fact that the profit gain dominates the demand forecasting error, from a different perspective by considering noisy demand functions. In these tests, we consider that the true estimated demand is perturbed by either an additive noise or a multiplicative noise. In other words, we assume that we estimate the demand from actual data as a deterministic function, and add some noise to test the robustness of our approach. We consider two common cases:

1. *Additive noise.* We assume that the demand d_t has a random additive noise that is uniformly distributed between −500 and 500. Note that the uniform case is the worst case, as more concentrated distributions will improve the results of the test. Note also that the average demand for this item (from the data we used in our case study) is around 1,000, so adding a noise between −500 and 500 is a good test. In this case, we apply the optimal prices suggested by solving our model with the deterministic demand and for each sample, we compute the profit gain. The results are summarized in Table 6 and Figure 7(a).
2. *Multiplicative noise.* We assume that the demand d_t has a random multiplicative noise that is uniformly distributed between 0.8 and 1.2. It means that we have between −20% and plus 20% of demand variation with respect to the forecast. In this case, we apply the optimal prices suggested by solving our model with the deterministic demand, and for each sample, we compute the profit gain. The results are summarized in Table 7 and Figure 7(b).

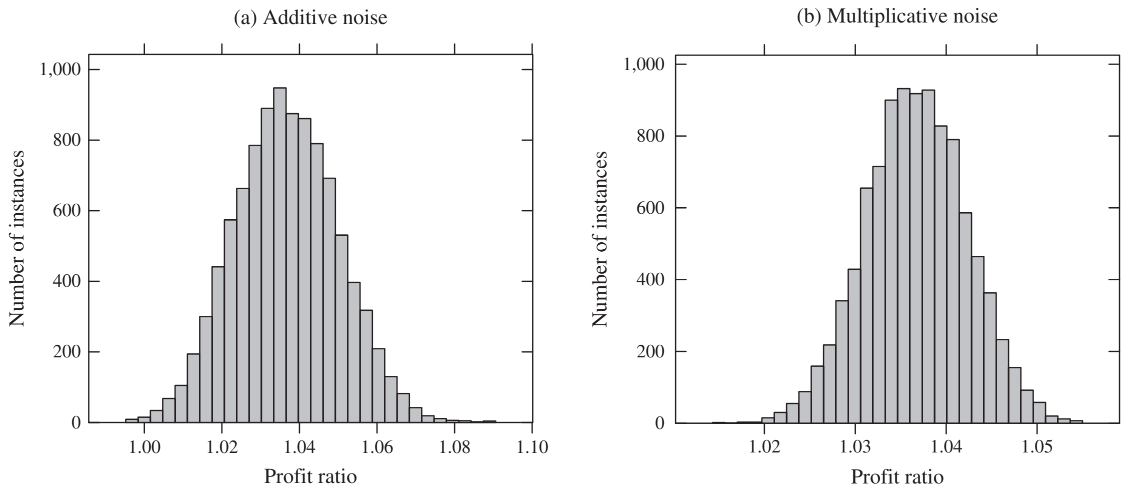
In both cases, one can see that the profit gain dominates the demand forecasting error. This again suggests that our approach is robust with respect to demand forecasting errors.

In many important settings, promotions are a key instrument for driving sales and profit. 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 general classes of demand

Table 6. Profit gain (in %) for randomly perturbed demands with additive noise (using 10,000 independent samples)

Nominal	Minimum	25th percentile	Median	75th percentile	Maximum
3.66	0.991	2.78	3.65	4.56	9.48

Figure 7. Histogram of profit ratios for 10,000 samples of perturbed demands with (a) additive and (b) multiplicative noise



functions depending on whether past prices have a multiplicative or an additive effect on current demand. These functions model the post-promotion dip effect that has been observed empirically, and can be 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 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 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. Finally, we demonstrate the robustness of our approach with respect to demand forecasting errors.

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, nonperishable products exhibit longer memory in the sense that the sales are affected not only by the current price but also by past prices. This observation validates the hypothesis that demand has a post-promotion dip effect for certain items. We test our approach for solving the POP, 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 3.5% 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.

Table 7. Profit gain (in %) for randomly perturbed demands with multiplicative noise (using 10,000 independent samples)

Nominal	Minimum	25th percentile	Median	75th percentile	Maximum
3.66	1.78	3.29	3.67	4.06	5.74

Acknowledgments

The authors thank the department editors, as well as the anonymous associate editor and reviewers for all their helpful comments and suggestions that helped to improve the content and exposition of this paper. The authors thank Oracle Corporation for their generous support through an External Research Office grant that made this work possible. A special thanks to Jeremy Kalas for his help in the computational parts of this work. The authors thank Steve Jeffreys from Oracle Retail for his suggestions and support. In addition, the authors thank Andrew Vakhutinsky and Peter

Gaidarev formerly from Oracle Retail for their valuable feedback in this work. The authors also thank Steve Graves, Adam Mersereau, Olivier Rubel, Paat Rusmevichientong, Chung-Piaw Teo and Huseyin Topaloglu for their insightful comments that helped to improve this paper. Finally, the authors thank the participants of the *2013 London Business School Collaborative Academic/Practitioner Workshop on Operational Innovation* for their valuable feedback and discussions.

Endnotes

[1] Private communication with Oracle executives.

[2] For confidentiality reasons, the actual name of the brand cannot be revealed.

[3] We also tested the hypothesis with nontruncated distributions such as Gaussian and exponential. In this case, the expectation formulation in Equation (17) yields the same optimal solution as in the nominal case, and is quite robust. However, the robust formulation provides a more conservative solution as expected.

References

Adams WP, Forrester RJ, Glover FW (2004) Comparisons and enhancement strategies for linearizing mixed 0-1 quadratic programs. *Discrete Optim.* 1(2):99–120.

Ahn Hs, Gümüs M, Kaminsky P (2007) Pricing and manufacturing decisions when demand is a function of prices in multiple periods. *Oper. Res.* 55(6):1039–1057.

Ailawadi KL, Gedenk K, Lutzky C, Neslin SA (2007) Decomposition of the sales impact of promotion-induced stockpiling. *J. Marketing Res.* 44(3):450–467.

Assunção JAL, Meyer RJ (1993) The rational effect of price promotions on sales and consumption. *Management Sci.* 39(5):517–535.

Bertsimas D, Shioda R (2009) Algorithm for cardinality-constrained quadratic optimization. *Comput. Optim. Appl.* 43(1):1–22.

Bertsimas D, Weismantel R (2005) *Optimization Over Integers*, Vol. 13 (Dynamic Ideas, Belmont, MA).

Bienstock D (1996) Computational study of a family of mixed-integer quadratic programming problems. *Math. Programming* 74(2):121–140.

Blattberg RC, Neslin SA (1990) *Sales Promotion: Concepts, Methods, and Strategies* (Prentice-Hall, Englewood Cliffs, NJ).

Buchbinder N, Feldman M, Naor JS, Schwartz R (2012) A tight linear time (1/2)-approximation for unconstrained submodular maximization. *Proc. 53rd Annual IEEE Sympos. Foundations Comput. Sci., FOCS '12* (IEEE Computer Society, Washington, DC), 649–658.

Cachon GP, Lariviere MA (2005) Supply chain coordination with revenue-sharing contracts: Strengths and limitations. *Management Sci.* 51(1):30–44.

Campo K, Gijsbrechts E, Nisol P (2000) Towards understanding consumer response to stock-outs. *J. Retailing* 76(2):219–242.

Caro F, Gallien J (2012) Clearance pricing optimization for a fast-fashion retailer. *Oper. Res.* 60(6):1404–1422.

Chaovalitwongse W, Pardalos PM, Prokopyev OA (2004) A new linearization technique for multi-quadratic 0–1 programming problems. *Oper. Res. Lett.* 32(6):517–522.

Chen X, Hu P, Shum S, Zhang Y (2014) Dynamic stochastic inventory management with reference price effects. *Oper. Sci.* 64(6): 1529–1536

Cohen MC, Gupta S, Kalas J, Perakis G (2016) An efficient algorithm for dynamic pricing using a graphical representation. SSRN: https://ssrn.com/abstract=2772231.

Cooper LG, Baron P, Levy W, Swisher M, Gogos P (1999) Promocast: A new forecasting method for promotion planning. *Marketing Sci.* 18(3):301–316.

Corsten D, Gruen T (2004) Stock-outs cause walkouts. *Harvard Bus. Rev.* 82(5):26–28.

Feige U, Mirrokni VS, Vondrak J (2011) Maximizing non-monotone submodular functions. *SIAM J. Comput.* 40(4):1133–1153.

Fibich G, Gavious A, Lowengart O (2003) Explicit solutions of optimization models and differential games with nonsmooth (asymmetric) reference-price effects. *Oper. Res.* 51(5):721–734.

Fletcher R, Leyffer S (1994) Solving mixed integer nonlinear programs by outer approximation. *Math. Programming* 66(1–3): 327–349.

Foekens EW, Leeflang SHP, Wittink DR (1998) Varying parameter models to accommodate dynamic promotion effects. *J. Econometrics* 89(1):249–268.

Gedenk K, Neslin SA, Ailawadi KL (2006) Sales promotion. *Retailing in the* 21*st Century* (Springer, Berlin), 345–359.

Grossmann IE (2002) Review of nonlinear mixed-integer and disjunctive programming techniques. *Optim. Engrg.* 3(3):227–252.

Gupta S (1988) Impact of sales promotions on when, what, and how much to buy. *J. Marketing Res.* 25: 342–355.

Heerde HJV, Leeflang PSH, Wittink DR (2000) The estimation of pre- and post-promotion dips with store-level scanner data. *J. Marketing Res.* 37(3):383–395.

Hemmecke R, Köppe M, Lee J, Weismantel R (2010) Nonlinear integer programming. 50 *Years of Integer Programming* 1958–2008 (Springer), 561–618.

Herer YT, Tzur M (2001) The dynamic transshipment problem. *Naval Res. Logist.* 48(5):386–408.

Kopalle PK, Rao AG, Assuncao JL (1996) Asymmetric reference price effects and dynamic pricing policies. *Marketing Sci.* 15(1):60–85.

Macé S, Neslin SA (2004) The determinants of pre- and postpromotion dips in sales of frequently purchased goods. *J. Marketing Res.* 41(3):339–350.

McCormick ST (2005) Submodular function minimization. Aardal K, Nemhauser GL, Weismantel R, eds. *Discrete Optimization.* Handbooks in Operations Research and Management Science, Vol. 12 (Elsevier, Amsterdam), 321–391.

Mela CF, Jedidi K, Bowman D (1998) The long-term impact of promotions on consumer stockpiling behavior. *J. Marketing Res.* 35(2):250–262.

Nemhauser GL, Wolsey LA (1988) *Integer and Combinatorial Optimization*, Vol. 18 (John Wiley & Sons, New York).

Nemhauser GL, Wolsey LA, Fisher ML (1978) An analysis of approximations for maximizing submodular set functions. *Math. Programming* 14(1):265–294.

Nielsen (2014) Cracking the Trade Promotion Code. *nielsen.com* (October 2, 2014), http://www.nielsen.com/us/en/insights/news/2014/cracking-the-trade-promotion-code.html.

Popescu I, Wu Y (2007) Dynamic pricing strategies with reference effects. *Oper. Res.* 55(1):98–112.

Porteus EL (1990) Stochastic inventory theory. Heyman DP, Sobel MJ, eds. *Stochastic Models.* Handbooks in Operations Research and Management Science, Vol. 2 (North-Holland, Amsterdam), 605–652.

Schrijver A (2000) A combinatorial algorithm minimizing submodular functions in strongly polynomial time. *J. Combinatorial Theory, Ser. B* 80(2):346–355.

Sherali HD, Adams WP (1998) *A Reformulation-Linearization Technique for Solving Discrete and Continuous Nonconvex Problems*, Vol. 31 (Springer, Berlin).

Su X (2010) Intertemporal pricing and consumer stockpiling. *Oper. Res.* 58(4-part-2):1133–1147.

Subramanian S, Sherali HD (2010) A fractional programming approach for retail category price optimization. *J. Global Optim.* 48(2):263–277.

Talluri KT, van Ryzin GJ (2005) *The Theory and Practice of Revenue Management* (Springer, New York).

Van Donselaar K, Van Woensel T, Broekmeulen R, Fransoo J (2006) Inventory control of perishables in supermarkets. *Internat. J. Production Econom.* 104(2):462–472.

Villas-Boas JM (1995) Models of competitive price promotions: Some empirical evidence from the coffee and saltine crackers markets. *J. Econom. Management Strategy* 4(1):85–107.

Zhao W, Zheng YS (2000) Optimal dynamic pricing for perishable assets with nonhomogeneous demand. *Management Sci.* 46(3): 375–388.

Maxime C. Cohen is an assistant professor of information, operations and management sciences at Stern School of Business, New York University. His research lies at the intersection of data analytics and operations management. He is interested in pricing applications such as retail, sustainability, airline, and cloud services.

Ngai-Hang Zachary Leung is an assistant professor in the Department of Management Sciences at the College of Business of the City University of Hong Kong. His current research interests include inventory management, revenue management, and healthcare operations.

Kiran Panchamgam is a principal scientist in the Retail Science R&D group at Oracle Inc, Burlington, MA. He has strong expertise (and also pending patents) in developing analytical solutions using data mining, machine learning, econometric and optimization techniques for business problems in retail.

Georgia Perakis is the William F. Pounds Professor of Management at the Sloan School of Management, Massachusetts Institute of Technology and is group head of the Operations Management group at MIT Sloan. In her research, she investigates how analytics can benefit complex real-world problems in pricing, revenue management, supply chain, energy and transportation applications. Perakis has received the CAREER Award from the NSF and the PECASE Award from the Office of the President on Science and Technology. She was recently elected as INFORMS Fellow, a recognition for lifetime contributions to the field.

Anthony Smith is Vice President of Science Services at Revionics, which provides price, markdown, and promotion optimization solutions for the retail industry. He has 30 years of experience in advanced analytic and science-based software, and for the past 12 years, has focused on retail data science and analytics, including demand forecasting, price optimization, supply chain and replenishment, and assortment space optimization. He holds an MS ECE from the University of Massachusetts, Amherst.