J Glob Optim (2010) 48:263–277
DOI 10.1007/s10898-009-9491-2

A fractional programming approach for retail category price optimization

Shivaram Subramanian · Hanif D. Sherali

Received: 21 March 2009 / Accepted: 5 November 2009 / Published online: 21 November 2009
© Springer Science+Business Media, LLC. 2009

Abstract We present a new mixed-integer programming (MIP) approach to study certain retail category pricing problems that arise in practice. The motivation for this research arises from the need to design innovative analytic retail optimization techniques at *Oracle Corporation* to not only predict the empirical effect of price changes on the overall sales and revenue of a category, but also to prescribe optimal dynamic pricing recommendations across a category or demand group. A multinomial logit nonlinear optimization model is developed, which is recast as a discrete, nonlinear fractional program (DNFP). The DNFP model employs a bi-level, predictive modeling framework to manage the empirical effects of price elasticity and competition on sales and revenue, and to maximize the gross-margin of the demand group, while satisfying certain practical side-constraints. This model is then transformed by using the Reformulation–Linearization Technique in tandem with a sequential bound-tightening scheme to recover an MIP formulation having a relatively tight underlying linear programming relaxation, which can be effectively solved by any commercial optimization software package. We present sample computational results using randomly generated instances of DNFP having different constraint settings and price range restrictions that are representative of common business requirements, and analyze the empirical effects of certain key modeling parameters. Our results indicate that the proposed retail price optimization methodology can be effectively deployed within practical retail category management applications for solving DNFP instances that typically occur in practice.

Keywords Multinomial logit model · Retail category pricing · Embedded demand model · Reformulation-Linearization Technique (RLT) · Mixed-integer programming · Fractional programming

S. Subramanian (✉)
Retail Analytic Foundation, Oracle, Suite 303, 25 1st Street, Cambridge, MA 02141, USA
e-mail: shiva.subramanian@oracle.com

H. D. Sherali
Grado Department of Industrial and Systems Engineering (0118), Virginia Polytechnic Institute and State University, Blacksburg, VA 24061, USA
e-mail: hanifs@vt.edu

1 Introduction

Product line pricing is an important business problem faced by retailers who employ dynamic pricing strategies to generate incremental revenue benefits throughout the year. Elmaghraby and Keskinocak (2003) review such dynamic pricing practices across industries and note that retailers, among others, have in increasing numbers begun to utilize decision support systems that leverage the large volume of detailed demand data to automate and optimize pricing recommendations. In particular, the statistical modeling of the price elasticity of items based on analyzing the effect of price changes of one product on its demand, or the demand for another product, is a well-researched area. For example, Reibstein and Gatignon (1984) estimate explicit pairwise inter-item cross-elasticity interactions to capture such effects and to determine optimal prices. Among models that implicitly capture inter-item interactions, the multinomial logit (MNL) model is a popular choice for discrete customer choice analysis (McFadden 1974; Ben-Akiva and Lerman 1985), and has come into prominence for product line pricing in the retail industry (Guadagni and Little 1983). In this context, the market share of an item is a consequence of its relative attraction with respect to other competing items (substitutes). A retailer would like to determine an optimal *category pricing strategy* to set prices for items in a given category (e.g., soups, cold cereals) for the next few weeks. These items are assumed to be substitutable in that they compete for the same customer dollar. However, unlike the cross-elasticity model, the MNL model generally cannot capture the market halo effects associated with complementary items (Train 1985). Several researchers have also endeavored to use such statistically calibrated models to generate optimal pricing recommendations. Hanson and Martin (1996) analyze the properties of the MNL-based optimization model by treating prices as continuous variables. They recognize the non-concavity of the profit objective function and present solution techniques for recovering a global optimal solution. Reibstein and Gatignon (1984) attempt to derive profit-maximizing prices for various categories of products based on certain statistically calibrated cross-elasticity models for predicting demand.

The regular pricing strategy employed by category managers is different from ‘markdown optimization problems’ that have been solved by retailers to determine optimal end-of-season pricing for clearance of inventory of short life-cycle products such as fashion products (e.g., designer apparel). On the other hand, the models described in this paper are more likely to be applicable to basic consumer items such as groceries. Also, business requirements generally involve multiple objectives. A popular approach is to maximize gross margin for the product category while also ensuring that the recommended prices result in category level sales and revenue values that meet preset targets. These targets can be conveniently quantified relative to the corresponding metrics at the original or current prices. Other constraints include price limits, as well as inter-item pricing rules that constrain the prices of pairs of items (e.g., a store brand soup should be priced 50c less than the corresponding national brand). In addition, the prices are required to be competitive with respect to competitor prices, whenever such data is available. In some situations, the number of price changes that can be recommended is also limited.

Whereas MNL models can be used to predict market shares based on relative utilities or attractions, the overall sales of the demand group (or category) itself does not change. To overcome this shortcoming, we assume that we are given a second demand model for predicting category-level demands as a function of some representative price of the items in the category. We briefly discuss the modeling impact of allowing multiple demand-subgroups whose items are substitutable only by items within the same subgroup. Other approaches such as the nested MNL model have also been considered for modeling multiple levels of customer choice (e.g., Silva-Risso and Ionova 2008). We assume that items are continually replenished, so that there is no inventory constraint. Consequently, the terms ‘sales’ and ‘demand’ are used interchangeably. Also, we do not analyze the sales impact of promotions or temporary price cuts within our model. Discussions on the merits and shortcomings of empirical demand models similar to the one adopted in this paper can be found in the marketing literature. For brevity, we highlight one such issue. A fundamental assumption regarding the *independence from irrelevant alternatives* (IIA) is a well-known drawback for MNL-based models, among others (McFadden 1974; Tse 1987). In essence, the IIA assumption in the retail category pricing context suggests that the ratio of the number of customers choosing any two substitutable items in the product line would be unaffected by the utilities of all other items in that set. On the other hand, the empirical results presented in Guadagni and Little (1983), based on an MNL model that was calibrated using market data for a consumer item (ground coffee), did not indicate significant or systemic estimation errors that were attributable to the IIA assumption.

The focus of the present work is to formulate a viable optimization model that can be applied to a wide variety of retail pricing contexts and to prescribe an effective solution methodology, assuming that we are given statistically calibrated demand models, and also that these demand models are compatible with the user. Our proposed model is based on real-world retail pricing problems faced at *Oracle Corporation*. Specifically, we derive a mixed-integer programming (MIP) formulation for optimizing an MNL model-based objective function in the presence of a discrete price ladder and other side-constraints, while also considering an extraneous competitive response to pricing, and we provide computational results on some realistic, randomly generated instances.

The remainder of this paper is organized as follows. In Sect. 2, we present an MNL-based margin optimization model for retail pricing that employs a dynamic category-level demand model and includes thresholds for sales and revenue goals. We propose a sequence of transformations using the RLT approach that enables us to initially pose the formulated model as an DNFP, and subsequently, as an equivalent MIP. In Sect. 3, we present some sample computational experience for the proposed solution methodology using realistic, synthetic data sets. Finally, Sect. 4 summarizes our findings and delineates specific directions for future work in this area of research.

2 MNL-based optimization model

In this section, we develop an MNL-based optimization model for category pricing in the retail industry. The scenario analyzed is as follows. Consider a retailer who has to set the baseline (or regular) price levels for some or all the active items in a given category for the next few months, as part of a merchandise planning process. The category manager has to make multiple, coordinated pricing decisions, proactively taking into account the impact of a price change on the sales of other items within the category, as well as any (extraneous) market response. Moreover, the recommended prices have to satisfy several category-level objectives such as profitability, sales, and revenue (e.g., to maximize gross margin while ensuring that the total sales and revenue are within 10% of the current value), and have to be selected from within a limited discrete price ladder (e.g., be within 20% of the current price and end with ‘9’ cents). In addition, items have to be priced relative to certain attributes such as brand type (e.g., a store brand tomato soup should be at least a dollar less than the price of the corresponding national brand), and quantity (e.g., a six-pack of diet-soda versus a two-liter bottle of diet-soda), among others. For a more detailed discussion of the requirements for category pricing, and dynamic pricing methods employed by retailers in general, we refer the reader to Elmaghraby and Keskinocak (2003).

Items can represent stock-keeping units (SKUs), product subclasses, or product classes within the category, depending on the level of aggregation in the merchandise hierarchy at which the analysis is performed by the category manager. For simplicity, we assume that we are optimizing prices of SKUs at the store-level of the location hierarchy, noting that the models in this paper can be readily extended to manage higher levels of aggregation (e.g., at the zonal level). Later in the paper, we briefly address more general situations faced by category managers such as the need to jointly optimize multiple categories that are inter-linked by pricing constraints and/or objectives, or manage several distinct subsets of substitutable items within the same category.

We first present an MNL nonlinear optimization model to represent this category pricing problem. Consider the following notation:

$n =$ number of substitutable items in the category or demand group.
$m =$ number of discrete price points or price levels per item.
$P =$ set of points in the price ladder.
$d_i =$ unit cost associated with item i.
$p_i^0 =$ initial (or original) price of item i.
$p_i =$ recommended price for item i (**principal decision variables**). Note that p_i can take on only one of certain pre-generated values (positive) $\bar{p}_{ij}, j = 1, \ldots, m$, in the *price ladder* for item i, $\forall\, i = 1, \ldots, n$.

$$z_{ij} = \begin{cases} 1 & \text{if } p_i = \bar{p}_{ij} \\ 0 & \text{otherwise}, \forall j = 1, .., m, \text{ for each } i = 1, .., n\cdot\ . \\ & (\textbf{Binary Decision Variables}) \end{cases}$$

$(l_i, u_i) =$ lower and upper bounds on the price for item i.
$p, l, u =$ vectors having respective components p_i, l_i, u_i.
$\theta =$ category-level sales (or demand) value (**function of the variables** $\mathbf{p}_i$, as given by (1b) below).
$\theta_0, R_0 =$ category-level (initial) sales, and revenue value, respectively, obtained by fixing prices of items at the initial prices p_i^0, $\forall i = 1, \ldots, n$.
$\Psi =$ price-elasticity parameter for category-level demand.
$U_i(p_i) =$ deterministic component of the utility of item i (**function of the variables** $\mathbf{p}_i$).
$\mu_i, \lambda_i =$ coefficients used in the utility expression for item i: $U_i(p_i) = \mu_i + \lambda_i p_i$.
$e^{U_i(p_i)} =$ measure of the attractiveness of item i to a consumer.
$\frac{e^{U_i(p_i)}}{\sum_{i=1}^n e^{U_i(p_i)}} =$ relative attractiveness measure of the MNL-predicted market share of item i within the category or demand group (Ben-Akiva and Lerman 1985).
$\alpha, \beta =$ coefficients used for setting category-level targets for sales, and revenue thresholds, respectively, which represent the respective threshold fractions of the initial sales and revenue values to be satisfied.
$K =$ number of inter-item constraints. These constraints manage relative "price movements" within item pairs (e.g., to maintain a consistent price-brand relationship).
$a_{i_k}, b_{j_k}, c_k =$ inter-item constraint coefficients (involving items i_k and j_k for each $k = 1, \ldots, K$).

$$v_{ij} \equiv e^{U_i(\bar{p}_{ij})} = e^{\mu_i + \lambda_i \bar{p}_{ij}},$$
$$r_{ij} \equiv v_{ij} \bar{p}_{ij}, \quad \text{and}$$
$$g_{ij} \equiv r_{ij} - v_{ij} d_i = v_{ij}(\bar{p}_{ij} - d_i), \quad \forall i = 1, \ldots, n,\ j = 1, \ldots, m.$$

We can then pose the category pricing problem as the discrete nonlinear fractional programming problem DNFP given below:

$$\textbf{DNFP}: \text{Maximize } \frac{\sum_{i=1}^{n} \sum_{j=1}^{m} g_{ij} z_{ij}}{\sum_{i=1}^{n} \sum_{j=1}^{m} v_{ij} z_{ij}} \theta \tag{1a}$$

subject to :

$$\theta = \theta_0 \left(\prod_{i=1}^{n} \frac{p_i}{p_i^0} \right)^{\frac{\psi}{n}} \tag{1b}$$

$$\theta \geq \alpha \theta_0 \tag{1c}$$

$$\sum_{i=1}^{n} \sum_{j=1}^{m} r_{ij} z_{ij} \theta \geq (\beta R_0) \sum_{i=1}^{n} \sum_{j=1}^{m} v_{ij} z_{ij} \tag{1d}$$

$$a_{i_k} p_{i_k} \leq b_{j_k} p_{j_k} + c_k, \quad \forall k = 1, \ldots, K \tag{1e}$$

$$\sum_{j=1}^{m} z_{ij}, = 1, \quad \forall i = 1, \ldots, n \tag{1f}$$

$$\sum_{j=1}^{m} \bar{p}_{ij} z_{ij} = p_i, \quad \forall i = 1, \ldots, n \tag{1g}$$

$$z \text{ binary.} \tag{1h}$$

Problem DNFP aims to determine the market share for each substitutive item in the category, while optimizing multiple objectives such as gross margin, sales, and revenue, and also considering the dynamic response of category-level sales to price changes and pricing rules. More specifically, it determines the price points in the price ladder for each item i that maximize the gross margin (1a) for the category. The objective function (1a) represents the sum of the market share-weighted item margins. The denominator in (1a) represents the sum of the MNL-based item-utility function values attained (based on the selected price points in the ladder), and is always positive-valued. The numerator in the ratio in (1a) represents the sum of the individual item profits realized $(\bar{p}_{ij} - d_i)$weighted by the corresponding utility function values (v_{ij}). If we assume that items are always priced more than their unit cost values ($l_i > d_i, \forall i = 1, \ldots, n$), then the numerator is also guaranteed to be positive-valued. Hence, the ratio yields the gross weighted profit per item, which is multiplied by the total category-level demand θ to compose (1a). The price effect on sales is captured using an empirical predictive model (1b) that attempts to explain the overall category sales as some nonlinear function of the geometric mean of the scaled prices of the items in the category based on an estimated value for the price elasticity parameter (Ψ) for category-level sales. Constraint (1c) ensures that the sales for the category does not fall below a given percentage of the original sales value, while Constraint (1d) imposes a similar restriction on revenue. Constraint (1e) captures simple inter-item pricing rules between items i_k and j_k, $\forall k = 1, \ldots, K$, as for example, rules that impose relative brand-price relationship, and Constraint (1f) selects a particular price from the designated price ladder for each item. Constraint (1g) identifies the price of each item based on the selected price point. (These relationships can be used to substitute the p_i-variables out of the model formulation.) Finally, Constraint (1h) enforces the binary logical restrictions on the z-variables.

Tawarmalani et al. (2002) present MIP reformulations for 0–1 hyperbolic programs (that resemble Problem DNFP for a fixed value of θ) and prescribe a global optimization scheme that employs a reformulation and linearization step at the nodes of a branch-and-bound tree to recover a global optimal solution. An alternative reformulation procedure and solution methodology is presented below.

We first apply the Charnes and Cooper transformation (1962) to (1a), while retaining the binary restriction on the z-variables to transform DNFP into an equivalent semi-continuous problem. Toward this end, denote

$$y \equiv \frac{\theta}{\sum_{i=1}^{n} \sum_{j=1}^{m} v_{ij} z_{ij}}, \tag{2a}$$

and let

$$x_{ij} \equiv y z_{ij}, \forall i, j. \tag{2b}$$

Note that $v_{ij} x_{ij}/\theta$ represents the market share of item i priced at $\bar{p}_{ij}$, while $v_{ij} x_{ij}$ represents the corresponding sales value. We next utilize the Reformulation-Linearization Technique (RLT) (Sherali and Adams 1999) that, in effect, linearizes the resulting bilinear terms $y z_{ij}$ in the objective function after applying (2), and provides an alternative representation for the semi-continuous variables x_{ij} (which take on values of y or 0 by (2b) and (1h)). Accordingly, we multiply (1a, 1d–1f) by ($y > 0$), but also retain (1f, 1h) as well as the original constraints (1e) to tighten the underlying relaxation, while eliminating the p_i-variables using (1g). For conveniently representing (1b), we take the logarithms (to the Naperian base e) in (1b) and introduce the relationship $w \equiv ln(\theta)$. Furthermore, for enabling the linearization of (2b), we impose lower and upper bounds y_{min} and y_{max} on the variable y, which are readily derived using (2a) and (1f, 1h), for example. This yields the following equivalent retail price optimization program (**RPO**), where (3b) represents (2a); (3c–3e) represent (1b, 1c) (where we have introduced an implied upper bound θ_{max} on θ in (3i) with accompanying bounds on w for subsequently linearizing (3c)); (3f) represents (1d); (3g, 3h) represent (1e); and the remaining constraints (3i–3m) represent (1f, 1h) along with the RLT linearization of (2b). (Also, recall that (1g) has been used to eliminate the p_i-variables throughout the problem.)

$$\textbf{RPO}: \text{Maximize} \sum_{i=1}^{n} \sum_{j=1}^{m} g_{ij} x_{ij} \tag{3a}$$

subject to:

$$\sum_{i=1}^{n} \sum_{j=1}^{m} v_{ij} x_{ij} = \theta \tag{3b}$$

$$w = \ln(\theta) \tag{3c}$$

$$w = \ln(\theta_0) + \frac{\psi}{n} \sum_{i=1}^{n} \sum_{j=1}^{m} \ln\left(\frac{\bar{p}_{ij}}{p_i^0}\right) z_{ij} \tag{3d}$$

$$\alpha\theta_0 \le \theta \le \theta_{\max}, \ln(\alpha\theta_0) \le w \le \ln(\theta_{\max}). \tag{3e}$$

$$\sum_{i=1}^{n} \sum_{j=1}^{m} r_{ij} x_{ij} \ge \beta R_0 \tag{3f}$$

$$a_{i_k} \sum_{l=1}^{m} \bar{p}_{i_k l} z_{i_k l} \leq b_{j_k} \sum_{l=1}^{m} \bar{p}_{j_k l} z_{j_k l} + c_k, \quad \forall k = 1, .., K \tag{3g}$$

$$a_{i_k} \sum_{l=1}^{m} \bar{p}_{i_k l} x_{i_k l} \leq b_{j_k} \sum_{l=1}^{m} \bar{p}_{j_k l} x_{j_k l} + c_k y, \quad \forall k = 1, .., K \tag{3h}$$

$$\sum_{j=1}^{m} z_{ij} = 1, \quad \forall i = 1, \ldots, n \tag{3i}$$

$$\sum_{j=1}^{m} x_{ij} = y, \quad \forall i = 1, \ldots, n \tag{3j}$$

$$x_{ij} \leq y_{\max} z_{ij}, \quad \forall i = 1, \ldots, n, \ j = 1, \ldots, m \tag{3k}$$

$$x_{ij} \geq y_{\min} z_{ij}, \quad \forall i = 1, \ldots, n, \ j = 1, \ldots, m \tag{3l}$$

$$y_{\min} \leq y \leq y_{\max}, z_{ij} \text{binary}, \quad \forall i, j. \tag{3m}$$

To facilitate the solution of Problem RPO using off-the-shelf MIP packages (such as CPLEX or GUROBI), we next linearize (3c), i.e., $\theta = e^w$, over the bounded rectangular region (3e) by restricting θ to lie below an S-segment piecewise linear inner approximation to the convex function e^w over the range $[\ln(\alpha\theta_0), ln(\theta_{\max})]$ (based on the restriction $\alpha\theta_0 \leq \theta \leq \theta_{\max}$), as well as to lie above some T tangential supports to this function sampled at points w_t, $t = 1, \ldots, T$, in the aforementioned range. This yields the following relationships that are used (henceforth) to replace Constraint (3c), where $\theta_t \equiv e^{w_t}$, $\forall t = 1, \ldots, T$.

$$\theta \leq \mathit{PiecewiseLinear}(w, S) \tag{4a}$$

$$\theta \geq \theta_t[1 + w - \log(\theta_t)], \quad \forall t = 1, \ldots, T, \tag{4b}$$

where *PiecewiseLinear* (w, S) denotes the aforementioned S-segment upper-bounding piecewise linear approximation to $\theta = e^w$. We used the piecewise linear modeling capability available within ILOG's CONCERT modeling language for our computational experiments, where the resultant MIP model generated by CONCERT employs S auxiliary binary variables for the internal representation of *PiecewiseLinear* (w, S). We report on the effect of experimenting with different values of S in the next section.

Our preliminary empirical analysis showed that the resulting optimal value of θ is primarily driven by the upper-bounding approximation. Consequently, we simply set $T = 2$, by providing tangential supports only at the end-points, $\ln(\alpha\theta_0)$ and $\ln(\theta_{\max})$, respectively. However, in practice, it is not guaranteed that the upper bounding approximation will always be active at optimality. Managers often add additional competitive price-matching constraints that tend to drive sales values of some items down if the corresponding competitor price is sufficiently high. Unprofitable items tend to have a similar effect on sales because of the margin-based objective (1a). On the other hand, some retailers may want to retain a limited inventory of "loss leaders" among these unprofitable items in their display shelves to induce in-store traffic. It is possible that a tradeoff resulting from a combination of such conditions can yield an optimal value of θ that lies strictly below the upper bounding functional.

In lieu of the piecewise linear approximation for θ, we can design an iterative process to determine a value of θ at which the corresponding (locally) optimal price values (nearly) satisfy (1b). In particular, when the structure of the empirical predictive model for category-level sales increases in complexity, such an iterative scheme may be the only practical choice available. Our empirical analysis based on the category-level sales model (1b) indicated that the foregoing MIP formulation derived from the prescribed piecewise linear approximation for θ was computationally adequate for obtaining relatively quick, good quality solutions for Problem RPO via the commercial software package CPLEX (version 10.0). Hence, we do not pursue such an iterative approach further.

Another important implementation issue is to derive a tight, valid bounding interval $[y_{\min}, y_{\max}]$ for the variable y. A simple way to determine an upper (lower) bound on y is via an optimal solution to the LP relaxation of Problem RPO, but replacing the objective function with one that maximizes (minimizes) y. However, rather than relying solely on feasibility considerations, it would be advantageous to derive these bounds while also taking into account the effect of the objective function (3a). Hence, we heuristically obtain a good quality feasible solution to Problem RPO having an objective function value ν (e.g., using the node-zero analysis embedded within CPLEX), and we then impose the valid inequality:

$$\sum_{i=1}^{n}\sum_{j=1}^{m} g_{ij}x_{ij} \geq \nu \tag{5}$$

within (3), while maximizing (or minimizing) y. A complementary approach to determine a potentially tight upper bound value for y is to employ the arithmetic mean–geometric mean inequality on the relationship (2a) to get, noting (3i, 3m) and the definition of v_{ij}:

$$y = \frac{\theta/n}{\frac{1}{n}\sum_{i=1}^{n}\sum_{j=1}^{m} v_{ij}z_{ij}} \leq \frac{\theta/n}{\prod_{i=1}^{n}\left[\sum_{j=1}^{m} v_{ij}z_{ij}\right]^{\frac{1}{n}}} = \frac{\theta/n}{\left[\prod_{i=1}^{n} e^{\sum_{j=1}^{m} U_i(\bar{p}_{ij})z_{ij}}\right]^{\frac{1}{n}}} \tag{6}$$

and accordingly maximize the logarithm of the right-hand side of (6), i.e., noting (3c), we

$$\text{maximize}\left\{w - \frac{1}{n}\sum_{i=1}^{n}\sum_{j=1}^{m} U_i(\bar{p}_{ij})z_{ij} - \ln(n)\right\}, \tag{7}$$

subject to the constraints of Problem RPO. We use the most restrictive resultant interval $[y_{\min}, y_{\max}]$ obtained from these approaches to generate Model RPO.

Remark 1 The underlying MNL model in (1) assumes that all the items in the demand group are substitutes. However, we frequently encounter customer data sets in practice that consist of several isolated demand subgroups (sub-categories), each of which contains items that are substitutable only by items within that sub-category. Allowing each such sub-category to be governed by its own statistically calibrated item-level MNL model, as well as its own sub-category-level demand model, helps improve the empirical performance of the overall predictive modeling framework. On the other hand, the items are required to have their prices jointly optimized across the group whenever items across subgroups are linked via the constraints of Problem DNFP. Given the index set G of such subgroups, it can be shown that the following generalization (8a, 8b), where we substitute

$$y_h = \frac{\theta_h}{\sum_{i=1}^{n}\sum_{j=1}^{m} v_{ij}^{h}z_{ij}}, \quad \forall h = 1, \ldots, |G| \tag{8a}$$

and let

$$x_{ij}^{h} = y_h z_{ij}^{h}, \quad \forall h, i, j, \tag{8b}$$

can be employed to derive a linear objective function $Z(G)$, given by:

$$\text{Maximize } Z(G) = \sum_{h=1}^{|G|} \sum_{i=1}^{n} \sum_{j=1}^{m} g_{ij}^{h} x_{ij}^{h}, \tag{8c}$$

where the x-, z- and y-variables, and the associated auxiliary coefficients are now identified by additionally indexing them based on their corresponding demand subgroup (h). Likewise, the corresponding side-constraints can be linearized to recover an MIP formulation similar to (3). The resulting MIP formulation can also be used to determine globally optimal pricing recommendations across several demand groups inter-linked by the constraints of Problem DNFP. These "superset" problem instances can involve several thousand pricing decision variables and inter-item pricing constraints, perhaps occurring in a block-diagonal pattern. We recommend the study of this extension for future research.

3 Computational tests

We present computational results in this section using realistic, randomly generated instances of Problem RPO. These individual instances were chosen to be representative of those arising from typical business requirements that we encountered at *Oracle Corporation*, and do not include peculiar instances that occur infrequently in practice, and that need to incorporate special practical safeguards within an industrial application, a discussion of which is beyond the scope of this paper. In reality, the regular price of basic items is changed relatively infrequently by retailers (promotions and discounts are used to make tactical price changes). Furthermore, such price changes are typically restricted to within 10–30% of the current price. Another common business requirement is that we obtain quick and good quality feasible solutions to enable rapid what—if kinds of analyses, while also ensuring that the solution methodology is reasonably accurately sensitive to changes in specification of the input.

Data generation and preprocessing:

(a) Synthetic data sets were generated for varying sizes, initial prices, initial sales, costs, MNL utility coefficients, and constraint settings.
(b) The category level price elasticity coefficient (Ψ) was assumed to be -2.0.
(c) The price ladder was generated using tables similar to the one shown below, while allowing a maximum variation of 30% from the original price value (p^0), and with the price points ending with the "magic number" 9 (cents). For example, an item that has to be priced between \$45 and \$55 will have the following price points: 45.49, 45.99, 46.49, 46.99, 47.49, 47.99, 48.49, 48.99, 49.49, 49.99, 50.99, 51.99, 52.99, 53.99, and 54.99.

Ladder	Low value (\$)	High value (\$)	Increment (cents)
1	0.00	9.99	10
2	10.00	49.99	50
3	50.00	99.99	100

(d) The upper and lower bounds for θ were tightened and the piecewise linear interpolation was generated over the resultant interval using a value of $S = 9$ (based on 10 uniformly spaced sample points, including the end points). We experimented with using fewer interpolating points and present some comparative results for these instances.

(d) The initial values for $y_{\min}$ and $y_{\max}$ were obtained as described in Sect. 2. In all cases, the geometric mean-based value for $y_{\max}$ was smaller than that obtained by maximizing y, subject to the constraints of Model RPO.
(d) The value for $y_{\min}(y_{\max})$ was repeatedly tightened via LPs by minimizing (maximizing) y, while revising the constraints (3k)–(3m) based on the best lower and upper lower bounds currently available for y, until the improvement in the bounds between iterations was less than 5%. On average, this iterative procedure reduced the initial value of $(y_{\max} - y_{\min})$ by about 25%, and no more than two iterations were required in all cases.
(d) Whenever feasible, we employed a node-zero based analysis to generate (5) and incorporated this constraint within the foregoing sub-problems for minimizing and maximizing y so as to further restrict $y_{\min}$ and $y_{\max}$ based on optimality considerations. We present some computational results to assess the incremental value of accommodating (5) in this step.

All runs were made on a 3.0 GHz Pentium 4 processor PC having 2 GB of RAM. The LP and MIP solvers available in CPLEX 10.0 were used to solve the various optimization (sub-) problems (we did not have access to a more recent version of CPLEX at *Oracle Corporation*). A limit of 2,000 enumerated branch-and-bound nodes, and a relative optimality gap target of 5% were used.

Table 1 presents results for randomly generated instances of problem RPO for different values of n as listed in Column 1. The corresponding number of bound-feasible price points (averaged over the n items) and the number of inter-item constraints generated are listed in the second and third columns, respectively. The optimal objective function value obtained for the LP relaxation of RPO is given in the fourth column, followed by the best feasible MIP objective function value achieved, in Column 5. The CPU seconds consumed by the overall approach is noted in Column 6, along with the number of branch-and-bound nodes enumerated (Column 7) and the relative optimality gap at termination (Column 8).

Feasible solutions within the preset optimality tolerance of 5% were obtained for all but one instance using the proposed MIP formulation. The optimal objective function value for the LP relaxation of Problem RPO was always within 8% of the best achieved MIP value. The preset branch-and-bound node enumeration limit of 2,000 was reached only for the largest instance, when the optimality gap at termination was about 6.4%.

To examine the effect of (5) on the solution quality and efficiency via the derivation of tighter bounds on y, we set ν equal to the best objective value found in Table 1 and added (5)

Table 1 Computational results for RPO

n	Avg. m	K	LP	MIP	Time (s)	Nodes	Opt. gap (%)
5	36	5	9,066	8,655	1	13	4.5
10	20	30	14,572	14,247	1	13	2.2
20	29	80	33,467	31,787	6	115	4.7
50	23	150	70,889	68,086	15	59	4.1
100	22	300	145,382	143,469	45	40	1.3
200	24	600	283,522	270,150	260	100	4.8
500	28	1,500	783,622	728,660	1,685	1,000	4.9
1,000	28	3,000	1558,868	1462,010	7,020	2,000+	6.4

+ Branch-and-bound node enumeration limit reached while minimizing and maximizing y. Table 2 analyzes the incremental impact of (5), with respect to the values presented in Table 1, on the achieved solution quality for a relative optimality tolerance target of 5%. For comparative purposes, the last column in Table 2 lists the best optimality gap achieved after reaching the preset node limit for each instance when we used a relatively tight optimality tolerance of 0.01%. On average, we observed a significant increase (about 10%) in the value of $y_{\min}$, and a relatively smaller reduction (less than 1%) in $y_{\max}$.

As can be seen in Table 2, when we solved (3) using the tighter bounds for y, we were able to significantly reduce the optimality gap at termination for most instances in comparison with Table 1. The largest instance in Table 1 could now be solved to completion, yielding a solution within 0.2% of optimality. On the other hand, we observed relatively limited improvement in the 20-item instance in Table 2, and were unable to reduce the optimality gap below 4.2% even after setting the optimality tolerance to 0.01% and enumerating 2,000 branch-and-bound nodes. However, using this tighter tolerance, all the other instances were solved to near-optimality (within 1.4%).

Note that it is possible to iteratively improve upon any incumbent feasible solution to (3) by employing the best achieved objective function value within (5) and repeating the aforementioned procedure until we obtain a solution within the desired level of optimality tolerance. In practice, a 5% optimality gap is generally acceptable, given the uncertain nature of the demand model parameters, among other factors.

Next, we varied the value of S to analyze the overall effect of employing different degrees of approximation for the piecewise linear representation of θ. Table 3 presents computational results for $S \in \{1, 2, 4\}$ (The results in Table 1 pertain to $S = 9$).

The trend in the achieved optimality gap values across Tables 1 and 3 for a given instance of Problem RPO indicates a general improvement as S decreases. Furthermore, in most instances, a smaller value for S required relatively less CPU time to achieve a similar solution quality. For example, using a value of $S = 1$ enabled the largest instance to also be solved within the desired optimality tolerance (in fact, the achieved objective function value for all instances was within 3.1% of optimality), while consuming a fraction of the CPU time used for the corresponding runs made for larger values of S. However, it is not guaranteed that a higher MIP objective function value obtained by using a smaller value for S will always translate into an equally improved value in the actual gross margin (1a) using the exact value

Table 2 Computational results for RPO while using (5) to bound y

n	LP	MIP (3% optimality tolerance)	Time (s) (3% optimality tolerance)	Nodes (3% optimality tolerance)	Opt. gap (%) (3% optimality tolerance)	Opt. gap (0.01% tol.)
5	9,046	8,741	1	33	3.3	1.4
10	14,563	14,251	1	43	0.01	0.01
20	33,467	31,690	8	88	4.9	4.2
50	70,884	68,147	22	100	0.1	0.01
100	145,338	143,544	45	100	1.2	0.01
200	283,204	271,487	157	200	4.3	0.05
500	782,005	729,743	1,456	753	0.3	0.1
1,000	1,557,275	1,464,170	10,805	1,640	0.2	0.1

Table 3 Computational results for RPO, for $S = 4$, 2, and 1

n	$S = 4$		$S = 2$		$S = 1$	
	Time (s)	Opt. gap (%)	Time (s)	Opt. gap (%)	Time (s)	Opt. gap (%)
5	1	3.0	1	1.9	1	3.1
10	1	2.0	1	1.5	1	0.4
20	12	4.0	7	3.5	2	1.3
50	17	3.8	10	3.1	3	0.3
100	56	1.1	54	0.8	10	0.3
200	358	4.2	253	3.2	34	1.0
500	1,678	4.9	1,598	4.9	371	0.5
1,000	6,960	6.0	6,480	5.5	3,900	0.2

Table 4 Relative over-estimation (error) in category level sales for various S

n	$S = 9$ (%)	$S = 4$ (%)	$S = 2$ (%)	$S = 1$ (%)
5	0.02	0.59	1.92	1.17
10	0	0.09	0.7	2.0
20	0.03	0.44	1.47	4.36
50	0.03	0.18	1.26	3.77
100	0.05	0.21	0.51	1.21
200	0.07	0.42	1.51	4.15
500	0.08	0	1.82	7.26
1,000	0.04	0.35	0.85	7.19

of θ via (1b) for the prices obtained. Table 4 tracks the relative error between the piecewise linear estimate of θ (which was active at optimality in all our test instances) and the true category-level sales (given by (1b)) for all instances presented in Tables 1 and 3, while Table 5 compares the trends in the recalculated (actual) gross margin values for these instances. The first column in Table 5 gives the actual gross margin dollars (i.e., the numerical objective function value) for $S = 9$, and the remaining columns tabulate the relative percentage increase or decrease in the gross margin values for other values of S, with respect to the values in the first column.

The results presented in Table 4 indicate that the choice of $S = 9$ resulted in a relative error of less than 0.08% for all instances, whereas using $S = 4$ limited the corresponding error to within 0.6%. A value of $S = 2$ corresponds to a two-segment piecewise linear approximation for θ, and this increased the relative error to between 0.7 and 2%. Setting a value of $S = 1$ essentially seeks to represent θ as a convex combination of its minimum and maximum values, thereby obviating the need for any additional binary variables to model the nonlinearity of θ as a function of w. The corresponding over-estimation in the category sales value was relatively quite high in this case, and ranged between 1.1 and 7.3%.

Any apparent or real improvement in the gross-margin obtained by reducing the value of S has to be weighed against any possible degradation in the achieved values of the secondary objectives specified via the side-constraints. For example, when S was reduced from 9 to 4, we observed that the true gross margin value improved or remained the same in six of the eight instances and yielded a relatively favorable tradeoff as far as the primary objective is

Table 5 Optimal gross margin (1a) for various values of S

N	$S=9$	$S=4$ (%)	$S=2$ (%)	$S=1$ (%)
5	8,653	0.94	0.64	0.44
10	14,247	−0.04	−0.01	−0.11
20	31,776	0.35	−0.21	−0.88
50	68,064	0.13	−0.27	0.03
100	143,392	0	0	−0.15
200	269,967	0.31	0.12	−0.13
500	728,095	−1.1	−0.02	−0.35
1,000	1,461,380	0.18	0	−0.94

concerned, while a further reduction of the value of S to 2 resulted in a drop in the actual gross margin value in four instances (see Table 5). Finally, when S was reduced to 1, the improved MIP objective function values did not translate into a true improvement in most (six) of the instances, with only a marginal improvement being observed in the other two instances. Whereas the values of $S=2$ and $S=1$ resulted in either an improved gross margin, or a relative reduction in gross margin value of less than 1% for all instances when compared to the corresponding values for $S=9$, we observed that the corresponding errors in any active side-constraints pertaining to sales, revenue, and CPI thresholds were relatively high (these results are not displayed for brevity). For example, the overestimation of the category sales value given in Table 4 is a direct measure of the achieved error in Constraint (1b). On the other hand, it is possible that a simple linear interpolation model for θ (using $S=1$) is acceptable for rapid 'ball-park' profitability assessments for typical real-life applications, especially when the user is mainly interested in getting a feel for the financial impact of allowing a limited number of pricing tweaks.

Finally, to assess the empirical effectiveness of the proposed approach on the average, we tested the overall procedure on ten randomized instances for each of the problem sizes shown in Table 1 (using a more recent version of CPLEX (11.2) that became temporarily available to us at *Oracle Corporation*). To generate these instances, we randomly varied the item costs, original prices, initial sales values, the category elasticity value, and utility coefficients, such that the corresponding LP-relaxation of (3) was feasible. We chose a value of $S = 4$, based on the discussion of the results in Table 4 in the previous paragraph, and without employing (5) to improve the bound on y. We imposed an LP subproblem enumeration limit of 2,000 if a feasible solution was at hand at that stage of the branch-and-bound analysis. Otherwise, we stopped at the first feasible integer solution beyond the 2,000-subproblem node limit in concert with an absolute subproblem enumeration limit of 10,000. The desired (relative) optimality tolerance was set at 0.01%. Each row in Table 6 presents the average values for the corresponding problem size in the following order from left to right: number of LP subproblems analyzed, CPU seconds consumed, index of the subproblem at which the first feasible integer solution was obtained, achieved optimality gap, the achieved error in the sales value, and finally, the percentage of instances that reached the 2,000-node enumeration limit. These average statistics only include instances for which a feasible integer solution was available at the end of the procedure. In all but one 1,000-item instance, a feasible integer solution was found before reaching the absolute node limit, and therefore, the last row of Table 6 reports values averaged over the nine instances for which a feasible solution was actually obtained for that problem size.

Table 6 Average computational effectiveness analysis of (3) using CPLEX 11.2

n	Nodes	Time (s)	First feasible node index	Opt. gap (%)	Error (%)	2,000-Node limit (%)
5	43	0.07	16	0.01	0.14	0
10	32	0.08	189	0.01	0.12	0
20	196	3.3	43	0.01	0.09	0
50	536	32	75	0.01	0.11	0
100	1,487	193	109	0.03	0.09	20
200	610	291	55	0.04	0.12	10
500	2,000	4,632	690	0.79	0.12	100
1,000	4,252	18,127	3,514	2.21	0.13	100

We can observe from Table 6 that, on the average, all instances having up to 500 items in the category could be solved to within 1% of optimality before reaching the first (2,000-node) enumeration limit. For small problem sizes (50 items or less), we were able to find the globally optimal solution in every single instance, while for medium sized instances (200 items or less), we were able to recover the global optimum in all but three instances. Furthermore, in all but one instance corresponding to the largest (1,000-item) data set, the achieved objective function value was within 2.21% of optimality on average. On the other hand, for the largest problem sizes (500 and 1,000 items), it was observed that most of the computational effort was directed toward finding the first feasible solution, and a minimum of 2,000 subproblems were enumerated in every such instance. In the case of the 500-item instances, we always found an integer-feasible solution before reaching this limit and, consequently, we stopped the procedure at or before this limit, whereas in the majority of the 1,000-item instances, we could not find such a feasible solution before reaching the aforementioned limit, but we found a feasible solution (in all but one instance) after analyzing an average of 3,514 subproblem nodes. On the whole, for 79 of the 80 individual instances tested, it was observed that the objective function value corresponding to the first feasible solution was always within 7% of optimality. As far as the error in the category sales value is concerned, on average, we achieved an error of 0.14% or less for all problem sizes (with a relatively limited deviation from this mean value) while employing the chosen value of $S = 4$, which is well within the acceptable practical range of allowable error.

4 Conclusions

We have proposed MIP formulations to model certain retail pricing problems that arise within real-life analytic applications implemented at *Oracle Corporation*. A bi-level empirical predictive framework that includes a high-level price elasticity-based demand model, as well as a multinomial logit-based market-share prediction model at the lower level, was embedded within an optimization framework to formulate a discrete, nonlinear fractional program (DNFP), which was subsequently transformed into an equivalent MIP having a relatively tight underlying LP relaxation. Our empirical analysis using CPLEX 10.0 indicates that typical price-range restricted instances of Problem RPO that arise in practice, ranging up to a thousand items in size, can be iteratively solved to near-optimality by incrementally tightening the associated LP relaxation and re-optimizing the resultant MIP. Incremental improvements in solution quality and computational performance that can be achieved for such instances by controlling the degree of approximation of the underlying nonlinear demand model have to be weighed against the achieved level of error in the nonconvex constraints, such as those involving sales and revenue targets that are derived from the demand model. Empirical evidence suggests that an effective commercial optimization software package can handle most of the practical scenarios encountered within an industrial implementation of our proposed modeling approach.

For future research, we recommend the modeling and analysis of problems that consider alternative price-elasticity based demand structures occurring in multiple sub-groups (see Remark 1). The MNL model is a popular option for quantifying customer-choice in a variety of other industries (e.g., travel). It is possible that the mathematical models described in this paper, perhaps with some minor modifications, can be extended to solve similar problems in other industries. In the immediate retail context, category managers often have to consider additional objectives arising from competition, product discounts, and assortment decisions, which result in complex optimization models characterized by an increasing degree of nonconvexity and non-separability of the objective function. Moreover, incorporating the effects of uncertainty in the demand model parameters within the formulation can potentially improve the robustness of the price recommendations. Although this would further complicate the model, we expect that the effect of transforming such structures into more manageable mathematical forms as done herein for Model RPO will be fruitful, and our preliminary research work in this area has been encouraging.

Acknowledgments This research has been supported in part by the *National Science Foundation* under Grant Number CMMI-0552676. We also thank Dr. V. Ramakrishnan at the MIT Sloan School of Management, and S. Jeffreys and the team of marketing scientists at the Retail Analytic Foundation, Oracle Corporation, for their support and valuable discussions during this research work.

References

Ben-Akiva, M., Lerman, S.: Discrete Choice Analysis: Theory and Application to Travel Demand. MIT Press, Cambridge (1985)

Elmaghraby, W., Keskinocak, P.: Dynamic pricing in the presence of inventory considerations: research overview, current practices and future directions. Manage. Sci. **49**(10), 1287–1309 (2003)

Guadagni, P.M., Little, J.D.C.: A logit model of brand choice calibrated on scanner data. Mark. Sci. **11**, 372–385 (1983)

Hanson, W., Martin, K.: Optimizing multinomial logit profit functions. Manage. Sci. **42**(5), 992–1003 (1996)

McFadden, D. Conditional logit analysis of qualitative choice analysis. In: Zarembka, P. (ed.) Frontiers in Econometrics, Academic Press, New York (1974)

Reibstein, D.J., Gatignon, H.: Optimal product line pricing: the influence of elasticities and cross-elasticities. J. Mark. Res. **21**, 259–267 (1984)

Silva-Risso, J., Ionova, I.: A nested logit model of product and transaction-type choice for planning automakers' pricing and promotions. Mark. Sci. (to appear) (2008)

Sherali, H.D., Adams, W.P.: A Reformulation–Linearization Technique for Solving Discrete and Continuous Nonconvex Problems. Kluwer, Dordrecht (1999)

Tawarmalani, M., Ahmed, S., Sahinidis, N.V.: Global optimization of 0-1 Hyperbolic Programs. J. Global Optim. **24**, 385–417 (2002)

Train, K.E.: Qualitative Choice Analysis: Theory, Econometrics, and an Application to Automobile Demand. MIT Press, Cambridge (1985)

Tse, Y.K.: A diagnostic test for the multinomial logit model. J. Bus. Econ. Stat. **5**(2), 283–286 (1987)