

A retail store SKU promotions optimization model for category multi-period profit maximization

Shaohui Ma^{a,b,1},

Robert Fildes^b

^a School of Economics and Management, Jiangsu University of Science and Technology,
ZhenJiang, 212003, China

^b Lancaster Centre for Forecasting, Lancaster University, Lancaster, LA1 4YX, UK

¹ Corresponding author at: School of Economics and Management, Jiangsu University of Science and Technology, ZhenJiang, 212003, China. Tel.: +86 138 15179032. E-mail address: msh@tju.edu.cn (Shaohui Ma); r.fildes@lancaster.ac.uk (R. Fildes).

Abstract

Consumer promotions are an important element of competitive dynamics in retail markets and make a significant difference in the retailer's profits. But no study has so far included all the elements that are required to meet retail business objectives. We extend the existing literatures by considering all the basic requirements for a promotional Decision Support System (DSS): reliance on operational (store-level) data only, the ability to predict sales as a function of prices and the inclusion of other promotional variables affecting the category. The new model delivers an optimizing promotional schedule at Stock-Keeping-Unit (SKU) level which maximizes multi-period category level profit under the constraints of business rules typically applied in practice. We first develop a high dimensional distributed lag demand model which integrates both cross-SKU competitive promotion information and cross-period promotional influences. We estimate the model by proposing a two stage sign constrained regularization approach to ensure realistic promotional parameters. Based on the demand model, we then build a nonlinear integer programming model to maximize the retailer's category profits over a planning horizon under constraints that model important business rules. The output of the model provides optimized prices, display and feature advertising planning together with sales and profit forecasts. Empirical tests over a number of stores and categories using supermarket data suggest that our model generates accurate sales forecasts and increases category profits by approximately 17% and that including cross-item and cross-period effects is also valuable.

Keywords: OR in marketing; promotion optimization; demand forecasting; fast-moving consumer goods retailing.

1. Introduction

Grocery retailing is a highly competitive industry. Small improvements in operational decisions may change the competitive balance permitting the chain to survive and prosper. On average, a supermarket's margin is about only 1% of net sales (Bolton, Shankar and Montoya, 2007; Chapados et al. 2014). Consumer promotions are an important element of competitive dynamics in retail markets with retailers using a myriad of promotional techniques to attract consumers. Given the widespread use of retail promotions and the magnitude of the dollars spent on them, promotions planning has the potential to make a significant difference in the retailer's profits.

The promotions planning process is complex and challenging for retailers. First, the promotion of one product affects not only the demand for the focal product in the current promotion period; it may also affect the demand for the product in the periods that follow, or even the demand for other product in the same store. This means that a promotion could be profitable in terms of the focal product in the focal period, but at the cost of reducing the profits in the following periods or sacrificing profit arising from other products. The profitability of a promotional plan should therefore be carefully accounted for. Second, as the manufacturers often provide numerous temporal deals on certain product, determining the timing of the promotion is also critical to the solution (Tellis and Zufryden, 1995). Third, promotions are often constrained by a set of business rules specified by the store and/or product manufacturers. Example of business rules include prices chosen from a discrete set with a limit to the number of promotions allowed over the planning horizon time (Silva-Risso Bucklin and Morrison, 1999; Cohen et al., 2014). Finally, the problem is difficult because of its large scale (Levy et al., 2004; Huang, Fildes, and Soopramanien, 2014). A typical supermarket today is bigger than ever before, with many thousands of items leading to a large number of weekly promotion decisions that have to be made.

We propose a two-step approach that allows Fast-Moving Consumer Goods (FMCG) retailers to address the dynamic and interactive implications of promotion decisions, generate accurate forecasts of sales and profit at store Stock-Keeping-Unit (SKU) level, and to make efficient promotion plans to maximize the product category's long-term profitability. Specifically, we first use a high dimensional Autoregressive Distributed Lag (ADL) model to

capture each SKUs' promotional interactive effects and cross periods promotional influences; this is estimated by a two-stage sign constrained regularization approach. Then in the second step, to optimize the promotion planning, we build a nonlinear integer programming model to maximize retailer whole category multi-period profitability under constraints that model important business rules.

To gain evidence for the concept, we test our approach on 915 SKUs, from four categories in four retailing stores. The results of our experiments show that the proposed model fits the data well in terms of accurate demand forecasting, outperforming all other alternative models. In term of the cumulative profit obtained, the approach can increase profits by on average 16.7% compared to the actual implemented promotion schedule.

The remainder of the paper is structured as follows. In section two, we review the related studies and address their limitations. In Sections three and four we discuss methodological issues. Section five describes the data, introduces the experimental design and presents the empirical results. Section six discusses the findings, offering conclusions as to promotional planning in practice and topics for further academic research.

2. A Review of Promotional Planning Models for Optimization

2.1 Promotional effects

Sales promotions are an important area of research in the field of marketing. It is well known that promotions often result in large sales effects for a promoted item. However, this does not mean that the sales increase is truly beneficial for a retailer. The sales increase for a promoted item could come from other items, or from other time periods because of stockpiling (van Heerde, Leeflang, and Wittink, 2002).

Researchers have found that the majority of the promotional response stems from brand switching, the percentage of own-brand sales elasticity with respect to a particular promotion that is due to brand-switching elasticity is estimated at about 30%-45% in studies published after 2002 (e.g., Sun, Neslin, and Narasimhan, 2003; van Heerde, Gupta, and Wittink, 2003; van Heerde, Leeflang, and Wittink, 2004; Sun, 2005; Hübner, Kuhn & Kühn, 2016). Also it is

widely recognized that for certain products, promotions may have a promotion fatigue effect, i.e., consumers may buy additional units of a product during promotions for future consumption (stock piling behavior). Researchers have found that, on average, 10%-20% of the promotion bump is due to accelerating or “pulling forward” future purchases in the store (Mela et al. 1998, van Heerde et al. 2000, Mace and Neslin 2004, Ailawadi et al., 2007b).

These findings show that a retailer may not benefit from promotion because of brand switching within the store, nor if the sales increase borrows from other time periods. Thus, a promotions optimizer developed for retailers should consider both promotional cross-item and cross-period effects to maximize the profitability of an entire category (or even the store) profit over a period of time rather than that of individual items in one period. If these operational issues are relaxed, the problem is greatly simplified but potentially leads to inferior results that are less practical.

2.2 Promotion optimization

A retailer’s optimal promotional decisions for items within a category of frequently purchased consumer nondurable product are complex and multifaceted. The retailer must determine the optimal prices of the various items in the product category and also how they vary with promotional activities such as feature advertisements and special displays (Vilcassim and Chintagunta, 1995). According to the type of data required by the model, existing promotions optimization research studies can mainly be classified into two streams: optimization based on household panel data and optimization based on store data.

2.2.1 Promotion optimization based on household panel data

Due to the availability of household panel data in 1980s, marketing researchers have used choice models extensively to study the effect of marketing mix variables on household brand choice behavior (e.g., Guadagni and Little, 1983; Gupta, 1988). In the 1990s, researchers began to explore how household scanner panel data could be used to determine optimal prices. For example, Vilcassim and Chintagunta (1995) developed a static optimization model of the retailer's pricing strategy based on a response model of consumer store visits, category incidence and brand choice. Tellis and Zufryden. (1995) also developed a dynamic model that optimized both the timing and depth of price discounts along with a retailer's order quantity in response to consumer decisions on incidence, brand choice and

quantity over time.

Any promotional cross-item effects can be naturally considered in the choice modeling framework. By estimating the promotional impacts on household stocking and brand loyalty, the long-term promotional influences can also be captured. Models based on panel data can also help to better understand how the promotions affect consumer brand switching behavior. Though appealing in theory, there are several barriers that limit the applications of promotion optimization approaches based on panel data. First, the availability of panel data is usually low for retailers. Adequate panel data are not collected in all markets and product categories and may have small sample sizes in many cases where they are collected. Second, even if relevant data are available; the problem of sample size and the sample bias still is a concern. Third, choice model cannot provide good forecasts of the aggregate product sales which is a basic requirement of a promotional Decision Support System (DSS) (Montgomery, 2005). So these optimization models based on panel data are useful in order to theoretically compare the results for different promotional policies (and derive managerial insights), but are difficult to apply in real promotional decision support system at store level. In fact, Kunz and Crone (2014) have argued that for meaningful applications store level EPOS data must be used.

2.2.2 Promotion optimization based on store data

Store data are far more likely to be used by managers for decisions about promotions because of their wide availability (Bucklin and Gupta, 1999). Researches on promotions optimization based on store data have a long history. Reibstein and Gatignon (1984) performed a pricing analysis based on an estimated double-log aggregate sales response function. They examined the issue of how a retailer should determine optimally the prices of the various sizes of eggs (extra-large, large, and medium, etc.). Blattberg and Neslin (1990) showed that a model of this type could also be used to determine whether to display an item or not, or whether to include an item in the feature advertisement or not. Mulhern and Leone (1991) went further seeking to evaluate the profitability of price discounts from a cross-category perspective and studying optimal retailer pricing and promotion policies. Though all of these researches considered the important issue of promotional cross-item effects, they implemented at single period static optimization, worked at brand level and failed to consider long-term promotional influences.

In their Marketing Science practice price winner paper, Ailawadi et al. (2007a) presented a methodology which can quantify promotions impact on store profitability by considering

brand switching, stockpiling and halo effects. But the purpose of their study was to quantify the net unit and profit impact of each promotion, not to optimize the specific promotion schedule (e.g., timing and price cutting depth), nor to provide a practical SKU level promotional DSS tool for retailers in contrast to the aims of this paper. They assumed that promoted items in a category have the same switching percentage and all types of promotion have the same impact. In contrast this paper distinguishes the heterogeneity of promotional elasticities among items, the asymmetric interactions, and the impact of different promotion types leading to a specific promotions schedule among the SKUs in category forecasts of SKU sales.

In recent years, a number of researchers have focused on SKU level promotion optimization. For example, Natter et al. (2007) described a dynamic retail pricing decision-support system for an Austrian Do-It-Yourself (DIY) retailer. They considered cross-product promotion effect based on market basket analysis, but neglected the longer term promotional influences. Also their model aimed at item profit maximization instead of category/store profit maximization. Cohen et al. (2014) also introduced a promotions depth and timing optimization formulation and proposed a linear approximation to solve the problem efficiently while captures several important business requirements as constraints, but their model also looked for single-item profit maximization and failed to consider cross-product promotional effects. Ferreira, Lee and Simchi-Levi (2015) introduced a pricing decision support system for an online retailer. Their objective was category level single period profit maximization: this neglects dynamic promotional influences, and also they simplify the cross-product effects by considering only the impacts of category average price. In short, while there have been a number of studies that have been based on store level SKU data, no study has so far included all the elements that are required to meet retail business objectives.

2.3 Contribution

This research develops a SKU store level promotions planning optimization method based on store data. We extend the existing literatures by considering all the basic requirements for a promotional DSS (Montgomery, 2005; Kunz and Crone, 2014; Agrawal & Smith, 2013): (i) reliance on operational (store-level) data only, (ii) the ability to predict sales as a function of prices and other input variables including feature and display advertising, (iii)

the inclusion of the constraints arising from business rules typically applied in practice. These deliver an optimizing promotional schedule at SKU level which aims to maximize multi-period category level profit. In estimating this optimal promotions schedule we consider promotional cross-product effects, long-term promotional influences and seasonality. A number of these critical issues have typically been neglected in earlier studies focused on price-promotion optimization at SKU level, in particular forecast validation and comparison with other types of promotional forecasting models. Table 1 provides a summary of comparisons of this research with existing representative studies. It demonstrates this study to be the first that includes all the features necessary for an effective promotions DSS.

Table 1
The studies on retailing promotion optimization

Paper	Data	Planning level	Cross-product influences	Cross-period influences	Forecast validation	Business Rules
Mulhern & Leone(1991)	Panel	Brand	Yes	No	No	No
Tellis &Zufryden (1995)	Panel	Brand	Yes	Yes	No	No
Vilcassim & Chintagunta(1995)	Panel	Brand	Yes	No	No	No
Ailawadi et al. (2007a)	Store	Category	Yes	Yes	No	No
Natter et al.(2007)	Store	SKU	Yes	No	No	No
Ferreira et al.(2015)	Store	SKU	Yes	No	No	No
Cohen et al.(2014)	Store	SKU	No	Yes	No	Yes
This study	Store	SKU	Yes	Yes	Yes	Yes

3. Demand model

The fundamental problem in the promotion optimization problem is to relate past price and promotional changes to sales. As we discussed in the previous section, at least three sources of promotional effect should be considered in the model: the promotional effect on the demand of focal SKU in current period, the effect on the demand of focal SKU in later periods, and the effect on the demand of other SKUs in the same category in the current period. We model all of these effects of promotion on demand with a high dimensional autoregressive distributed lag model.

3.1 Autoregressive distributed lag model

To capture the promotional impacts on demand, and building on its strong forecasting performance as demonstrated in Huang et al. (2014), Ma, Fildes and Huang, (2016), we model the demand of SKU j in category k in week t as an autoregressive distributed lag (ADL)

model

$$\begin{aligned} \ln(Y_{kjt}) = & \eta_{kj} + \sum_{l=0}^1 \beta_{kjl}^j \mathbf{X}_{kjt-l} + \alpha_{kj}^j \ln(Y_{kj,t-1}) + \rho_{kj} \ln(\tilde{Y}_{kj,t-1}) \\ & + \sum_{i=1, i \neq j}^{n_k} [\beta_{ki}^j \mathbf{X}_{ki,t} + \alpha_{ki}^j \ln(Y_{ki,t-1})] + \sum_{c=1}^9 \sum_{l=0}^1 \delta_{kjc}^c C_{t-l}^c + \varepsilon_{kjt} \end{aligned} \quad (1)$$

where

$\ln(Y_{kjt})$ is the log sales of the focal product j in category k in week t ;

η_{kj} is the product j 's specific constant;

$\mathbf{X}_{ki,t} = [P_{ki,t}^R / P_{ki,t}, D_{ki,t}, F_{ki,t}]$ is a vector of promotional explanatory variables including the ratio of the regular price and the list price of the product i in category k in week t , an indicator variable for display (1 if product j is displayed, in week t ; 0 otherwise), and an indicator variable for feature advertising (1 if product j is featured, in week t ; 0 otherwise);

$\ln(\tilde{Y}_{kjt})$ is the log eight weeks moving average sales of the focal product j in category k in week t , which is employed here to capture any longer term trend;

C_{t-l}^c is the dummy variable for the c^{th} calendar event at week $t-l$. When $l=0$, the dummy variable represents the week of the calendar event, and the week before the event if $l=1$; c take the values from 1 to 9 representing all the calendar events including Halloween, Thanksgiving, Christmas, New Year's Day, President's Day, Easter, Memorial Day, 4th of July, and Labour Day.

The β_{kjl} , l equals 0 or 1, is a coefficient vector of the promotion multiplier, the α_{kj} is the multiplier for sale lag of product j in category k , δ_{cj} is the calendar multiplier for event c and the disturbance term is represented by ε_{kjt} .

The ADL model described in Eq. (1) considers the entire cross promotional effects from other SKUs. To distinguish with other models in the paper, we hereafter call the model as ADL-CROSS model.

Weekly sales of grocery products have markedly skewed distributions characterized by a few extremely high values resulting from deep price cuts (Rajum, 1995; Rinne and Geurts, 1988). The logarithmic transformation of the dependent variable at least approximately normalized the distribution of the dependent variable. The lags included for both sales and explanatory variables can capture the longer term promotional effects, e.g., consumers may stockpile products during promotions, leading to reduced demand following the promotion. We empirically find that the proposed ADL-CROSS model can improve forecasting accuracy

by between 30% and 40% over the model without lags.

The ADL-CROSS model (1) suffers from three problems we need to handle. First, the number of explanatory variables is close to or even larger than the sample size because of the inclusion of cross promotional explanatory variables from other SKUs in the category. Ordinary Least Squares is therefore not appropriate in this high-dimensional linear model. Second, with high dimensionality, important predictors can be highly correlated with some unimportant ones, and the maximum spurious correlation also grows with dimensionality (Fan and Lv, 2008). This results in the promotional explanatory variables from different SKUs being multicollinear which makes it difficult to distinguish their individual effects on the dependent variable. Third, SKU level sales data is much noisier than higher aggregation level (e.g., brand). The promotional effects, especially the cross SKU and cross period promotional effects are often too weak to measure, result in incorrect signs for the estimated promotional explanatory variables (Mullet, 1976). Incorrect signs may not be a big problem in forecasting applications, but they may be seriously misleading in applications which focus on promotions optimization. To deal with these problems, we propose a two-stage sign constrained regularization approach for the model's estimation which is described as following.

3.2 Sign constrained regularization

We separate the estimation process into two stages. At the first stage, we estimate the model with only the focus SKU's own promotional predictors as well as the calendar dummies by minimizing the following sign constrained L1 regularized least squares as follows:

$$\min_{\mathbf{a}} \sum_{t=1}^T \left\| \ln(Y_{kjt}) - \eta_{kj} - \alpha_{kj}^j \ln(Y_{kj,t-1}) - \rho_{kj} \ln(\tilde{Y}_{kj,t-1}) - \sum_{l=0}^1 \beta_{kjl}^j \mathbf{X}_{kj,t-l} - \sum_{c=1}^9 \sum_{l=0}^1 \delta_{kjc}^j C_{t-l}^c \right\|_2^2 + \lambda \|\mathbf{a}\|_1 \quad (2)$$

subject to $\beta_{kj0}^j \geq 0$ (promotional impacts in current period) and $\beta_{kj1}^j \leq 0$ (promotional stocking effects). $\|\cdot\|_1$ and $\|\cdot\|_2$ denote the l_1 and l_2 norms respectively;

$\mathbf{a} = [\eta_{kj}, \alpha_{kjl}^j, \beta_{kjl}^j, \delta_{kjc}^j, \rho_{kj}]$ is a coefficient vector including all the coefficients in the model to be minimized. λ is a nonnegative penalty parameter which determines the sparseness of \mathbf{a} . Setting $\lambda=0$ will reverse the problem to sign constrained least squares. In this empirical study the optimal value of λ for each SKU is determined by a 20-fold cross-validation..

Assuming the data in time window $[1, T]$ is used for model estimation, after the parameter estimation by the sign constrained L1 regularized least square in Eq.(2), we calculate the in-sample forecasts error $\hat{y}_{kj,1:T}$. At the second stage, we use the in-sample forecasts error $\hat{y}_{kj,1:T}$ from the first stage as the dependent variable, and use the promotional variables from the other SKUs in the same category as the focal SKU as the explanatory variables and estimate the remainder of the parameters in Eq.(1) by the second sign constrained L1 regularized least square,

$$\min_{\mathbf{b}} \sum_{t=1}^T \left\| \hat{y}_{kjt} - \sum_{i=1, i \neq j}^{n_k} \left[\alpha_{ki}^j \ln Y_{ki,t-1} + \beta_{ki}^j \mathbf{X}_{ki,t} \right] \right\|_2^2 + \lambda \|\mathbf{b}\|_1 \quad (3)$$

subject to $\beta_{ki}^j \leq 0, \forall i$, where $\mathbf{b} = [\alpha_{ki}^j, \beta_{ki}^j]$.

Penalized L1 likelihood methods have been successfully developed over recent decades to cope with high dimensionality. Penalized L1 regression is called LASSO by Tibshirani (1996) in the ordinary regression setting and has received much attention due to its convexity and the sparsity of its parameter space solutions. There has been much work in recent years, applying and generalizing the LASSO and L1-like penalties to a variety of problems (Tibshirani, 2011). By regularizing with an L1 penalty, we can generate sparse models and therefore cope with high dimensionality as in model (1). Using the proposed two stage estimation strategy, we separate explanatory variables into two information sets: own promotional information and cross-promotional information. This can avoid the regularizing estimator selecting correlated products' promotion variables instead of the focal SKU's own predictor because of aforementioned spurious correlation. By restricting the signs of the promotion coefficients, we can guarantee economic consistency in the model as well as further help the regularizing estimator select appropriate predictor variables (Lan et al., 2014).

3.3 Alternative demand models

3.3.1 ADL-OWN model

The ADL-OWN model represents the model using only the focal SKU's own predictors without considering cross-SKU promotional effects (Gür Ali, & Yaman, 2013; Williams et

al.,2014; Lang et al., 2015). This is a simplified low dimension version of ADL-CROSS described in Eq.(1) by neglecting all the promotional information from other SKUs. By comparing the performance of ADL-OWN and ADL-CROSS, we can infer the value of cross-item promotional information in forecasting and promotions planning.

3.3.2 Top-down with market share decomposition

Let Y_{kt} be the aggregate demand of category k at week t , and \bar{X}_{kt} be a vector of the promotional intensity indexes which are defined as the weighted average of the corresponding promotional variables, i.e., price, display and feature, across SKUs in the category. The weight is the weekly average sales in the calibration period. Then the demand at category level can also be modelled as an ADL model

$$\ln(Y_{kt}) = \eta_k + \alpha_k \ln(Y_{k,t-1}) + \rho_k \ln(\tilde{Y}_{k,t-1}) + \sum_{l=0}^1 \beta_{il}^k \bar{X}_{k,t-l} + \sum_{c=1}^9 \sum_{l=0}^1 \delta_{kcl} C_{t-l}^c + \varepsilon_{kt}, \quad (4)$$

with calendar event dummy variables, C_t included as before.

To decompose the aggregate category level demand to SKU level demand, we model the market share of SKU j in the category k with a market attraction model

$$A_{kjt} = \exp \left[\eta_{kj} + \alpha_{kj} \ln(M_{kj,t-1}) + \sum_{l=0}^1 \beta_{kjl} \mathbf{X}_{kj,t-l} + \varepsilon_{kjt} \right], \quad (5)$$

where A_{kjt} is the attraction of SKU j at time t , and $M_{kj,t-1}$ is the market share of SKU j in category k at week $t-1$. The market share of SKU j at week t is equal to its attraction relative to the sum of all attractions, that is,

$$M_{kjt} = \frac{A_{kjt}}{\sum_{j=1}^{n_k} A_{kjt}}. \quad (6)$$

Market share attraction models have been seen as useful tools for analyzing competitive structures (Cooper and Nakanishi, 1988; Cooper, 1993). The models can be used to infer cross-effects of marketing-mix variables, but one can also learn about the effects of the focal promotional actions while conditioning on competitive reactions (Fox, 2003). To estimate the parameters in the model (5), we first select one SKU in the category as the benchmark, and

then we take the natural logarithm of the ratios between the market shares of the reminder SKUs and the benchmark. This results in a system of the (n_k-1) -dimensional set of equations which is estimated by a seemingly unrelated regression (SUR) estimator (Zellner, 1962).

3.3.3 Promotional index model

Another model we considered is a SKU level promotional index model. Similar to the first model (1), again the demand of SKU j in sub-category k in week t is modeled with an autoregressive distributed lag (ADL) model, but in order to reduce the dimension of the explanatory variables, the promotional intensity indexes of the category are used as the predictors instead of SKU level high dimensional promotional variables

$$\begin{aligned} \ln(Y_{kjt}) = & \eta_{kj} + \alpha_{kj}^j \ln(Y_{kj,t-1}) + \sum_{l=0}^1 \beta_{kjl}^j \mathbf{X}_{kj,t-l} + \rho_{kj} \ln(\tilde{Y}_{kj,t-1}) \\ & + \sum_{l=0}^1 \beta_{kl}^j \bar{\mathbf{X}}_{i,t-l} + \sum_{c=1}^9 \sum_{l=0}^1 \delta_{kjl}^c C_{t-l}^c + \varepsilon_{kjt} \end{aligned} \quad (7)$$

All the promotional intensity indexes are calculated by weighted averaging the corresponding promotion values (discount, display and feature) across SKUs in a category. The weight is the weekly average sales of the SKU. That is, the larger the market share a SKU occupies in a category, the larger the weight it has in the calculation of promotional intensities. The model is estimated by OLS with robust estimators in the presence of heteroscedasticity. While the merit of this approach is that it is less computationally complex and easier to implement, it may suffer from the problem of loss of promotional information.

Table 2

A comparative summary of the alternative demand models

	ADL-OWN model	ADL-CROSS model	Top-down model	Promotional index model
Cross-item promotional effects	No	Yes	Yes	Yes
Asymmetric interactions	No	Yes	No	No
Heterogeneous impacts	No	Yes	No	No
Sign constrained	No	Yes	No	No

In order to have a better understanding of the substantial differences of the presented demand models, we summarize the most important properties of them in Table 2. Except for ADL-OWN model, all other models consider cross-item promotional effects. But both Top-down model and promotional index model simplified the cross-item effects by

neglecting asymmetric interactions between two SKUs and heterogeneous impacts from one SKU on others.

4. Optimization model

We consider the multi-periods promotion optimization problem for multiple SKUs in the same category. The objective is to maximize the total profits over some finite time horizons, whereas the decision variables are for each time period: whether and how to promote a SKU (display and feature advertising) and what price to set (i.e., the promotion depth).

We suppress the category subscript k to reduce notational clutter in the following. Let Ω_j be the discrete set of admissible prices for SKU j , then $G_j=|\Omega_j|$ represents the number of possible prices. It is common for many retailers to choose prices that end in .90 or 0.99 (i.e. \$1.90 or \$2.99: Gedenk and Sattler, 1999; Kunz and Crone, 2014). Our formulation of the discrete price set can easily meet this requirement in reality. Let p_{jg} represent the g th possible price in set Ω_j , where $g=1, \dots, G_j$. The first price in the set is regarded as the regular price and the others are regarded as price reductions. Define binary variables x_{jgt} such that $x_{jgt}=1$ if SKU j is assigned price p_{jg} at week t , and $x_{jgt}=0$ otherwise, for all $j=1, \dots, N$ where N is the total number of SKUs in the category. Also let C_{j1} be the unit cost of the SKU j . Then the gross margin when selling one unit of SKU j is

$$\sum_{g=1}^{G_j} x_{jgt} p_{jg} - C_{j1}.$$

The uncertainty in our model is given by Y_{jt} , a random variable representing the demand for SKU j at week t . Let \hat{Y}_{jt} be the expected demand for SKU j at week t which could be estimated with the proposed demand forecasting models described in Section 3. Define binary variables d_{jt} and f_{jt} to be the display and feature advertising indicator for SKU j in week t . Define binary variable l_{jt} to be the indicator for price tag changing, such that $l_{jt}=0$, if $x_{jgt}=x_{jg,t-1}$ for $\forall g$ (that means the price is unchanged in period t), otherwise $l_{jt}=1$; Let C_{j2} be the unit cost of the display of the SKU j , C_{j3} be the unit cost of the feature advertising of the SKU j , and C_{j4} be the unit cost of changing the price tag of the SKU j . Then if there are N SKUs in the category and H periods for promotions planning, the total profits the store can obtain from the category in the promotions planning period is

$$\text{Profits: } \max \sum_{t=1}^H \sum_{j=1}^N \left[\left(\sum_{g=1}^{G_j} x_{jgt} p_{jg} - C_{j1} \right) \hat{Y}_{jt} - C_{j2} d_{jt} - C_{j3} f_{jt} - C_{j4} l_{jt} \right] \quad (8)$$

The decision variables need to be set here are x_{jgt}, d_{jt}, f_{jt} all $\in \{0,1\}, \forall j, g, t$, and the objective is the maximization of the category profit over H periods. In addition, we need ensure that each SKU is assigned exactly one price corresponding to one possible value of the sum of all prices, that is

$$\sum_{g=1}^{G_j} x_{jgt} = 1 \quad \forall j, t. \quad (9)$$

We also need to consider various important real-world business requirements that should be satisfied. First, at a certain week t , the number of products concurrently under price reductions, in display or feature advertising should be constrained. Too many or too few products with price reductions in the same week may counteract the promotional effects or damage the store's price image. The positions for display or feature advertising are also limited resources for a retailing store. Let Lp_t and Up_t be the lower and upper bounds on the number of SKUs allowed to have a price reduction in the category in week t ; Ld_t and Ud_t be the lower and upper bounds on the number of SKUs allowed to be displayed in week t ; and Lf_t and Uf_t be the lower and upper bounds on the number of SKUs allowed to have feature advertising in week t , then we set

$$Lp_t \leq \sum_{j=1}^N \sum_{g=2}^{G_j} x_{jgt} \leq Up_t \quad \forall t; \text{ (SKUs in price reduction at } t) \quad (10)$$

$$Ld_t \leq \sum_{j=1}^N d_{jt} \leq Ud_t \quad \forall t; \text{ (SKUs in display at } t) \quad (11)$$

$$Lf_t \leq \sum_{j=1}^N f_{jt} \leq Uf_t \quad \forall t; \text{ (SKUs in feature advertising at } t) \quad (12)$$

Second, the frequency of the promotions for a product over time should also be limited. To have too many periods for a product in promotion would change the price image in consumers' mind. Let Mp_i (Md_i, Mf_i) be the maximum number of weeks SKU_i is allowed to have a price reduction (in display, in feature advertising) in the planning horizon H, then we set

$$\sum_{t=1}^T \sum_{g=2}^{G_j} x_{jgt} \leq Mp_j \quad \forall j; \text{ (periods in price reduction for SKU } j) \quad (13)$$

$$\sum_{t=1}^T f_{jt} \leq Mf_j \quad \forall j; (\text{ periods in display for SKU } j) \quad (14)$$

$$\sum_{t=1}^T d_{jt} \leq Md_j \quad \forall j; (\text{ periods in feature advertising for SKU } j) \quad (15)$$

The proposed multi-periods promotion optimization model described in (8)-(15) is a nonlinear integer programming problem and is in general hard to solve with traditional analytical methods. Thus, we resorted to a Genetic algorithm to derive the levels of price reduction, display and feature advertising desired for each SKU that maximize the category profit in the planning horizon (Scrucca, 2013). Genetic algorithms (Goldberg, 1989) are simulation-based, parallel-search algorithms that have been used in econometrics (Dorsey and Mayer, 1995; Liang and Wong, 2001), marketing (Venkatesan and Kumar, 2004; Venkatesan, Krishnan, & Kumar, 2004) and operational research (Nissen, 1995) to obtain optimal solutions when the complexity of the optimization function tends to be intractable and multi-dimensional. This method is not susceptible to local minima and it asymptotically converges with probability one to a globally optimal configuration. This feature allows the system to explore different regions of the solution space to find the global optimum. The algorithm, programmed in R, is described in this paper's supplementary material.

5. Empirical applications

5.1 Data

The empirical data comes from the IRI dataset (Bronnenberg et al., 2008)². The IRI dataset includes grocery and drug chain data from a sample of stores in 50 markets and 30 categories, involving approximately 25%-30% of the consumer packaged goods sales in a grocery store. This is weekly data by SKU and includes information on sales, price, features and displays. Based on the objectives of this research, the store scanner data from four medium-sized grocery stores in Chicago have been selected as the empirical sample. We apply the proposed demand forecasting and promotions optimization models to four product categories: milk, beer, mayo and yogurt, 915 SKUs in total. The data includes the weekly units sold, prices, displays and features for 280 weeks. The first 200 weeks were used to obtain model parameter estimates and the last 80 weeks for the validation and policy comparison. A sample series for one of the SKU is plotted in Fig. 1 Low-movement SKUs or

² All estimates and analyses in this paper based on Information Resources, Inc. data are by the author and not by Information Resources, Inc.

SKUs which may have been introduced or discontinued were excluded. Our criterion was that at least 80% of the weeks must have positive movement for the SKU. Kunz and Crone (2015) have shown the importance of including many SKUs in the optimization formulation.

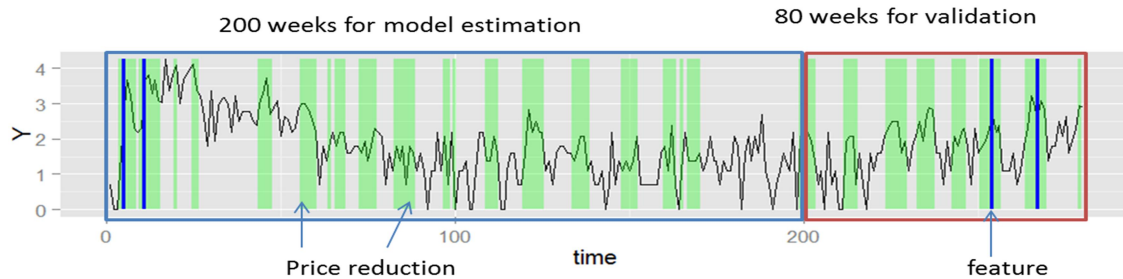


Fig.1. A sample series of SKU sales and promotions

Table 3 presents descriptive statistics of the data sample. For each store and each product category, the means and medians of units sold per week and percentages of weeks concerning promotional activities, including price reductions (more than 5 percent), displays and features are presented. It is clear that price reduction is the most frequent type of promotion across all the categories. Feature advertising is also frequently used in some categories in some store, such as yogurt in store 1 and beer in store 4. Display is used less frequently except for beer in store 4.

Table 3

Description statistics of the data sample

Store	category	Num of SKUs	Mean (Median) units sold per week	Percentages of weeks concerning promotional activities		
				Discounts	Displays	Features
1	Milk	36	127.83(41)	0.17	0.00	0.09
	Beer	72	6.37 (4)	0.26	0.04	0.08
	Mayo	10	6.11(4)	0.18	0.04	0.05
	Yogurt	84	37.07(27)	0.33	0.00	0.14
2	Milk	59	166.51 (102)	0.09	0.00	0.06
	Beer	140	9.97 (6)	0.07	0.02	0.06
	Mayo	19	25.28 (16)	0.08	0.02	0.04
	Yogurt	105	91.49 (66)	0.17	0.01	0.10
3	Milk	47	198.60 (56)	0.11	0.01	0.03
	Beer	10	8.47 (5)	0.11	0.03	0.05
	Mayo	21	41.89 (22)	0.13	0.04	0.02
	Yogurt	105	100.13 (67)	0.20	0.05	0.07
4	Milk	40	63.67 (23)	0.17	0.00	0.06
	Beer	98	12.92 (8)	0.29	0.26	0.12
	Mayo	17	16.37 (12)	0.21	0.03	0.08
	Yogurt	52	49.19 (37)	0.26	0.02	0.08

5.2 Estimation and forecasting validation for the demand models

We evaluate the forecasting performance of our models with ten rolling origins, which partially controls for any specific effects arising from a particular origin. Specifically, we estimate the model with a moving window of 200 weeks and then generate one week ahead forecasts for the next 8 weeks. We move the estimation window forward every eight weeks throughout the remaining sample period and re-estimate the model based on the updated data sets. The implement of the proposed two-stage sign constrained regularization approach is mainly based on R package GLMNET (Friedman, Hastie & Tibshirani, 2010). This is fully documented in the supplementary material to this article.

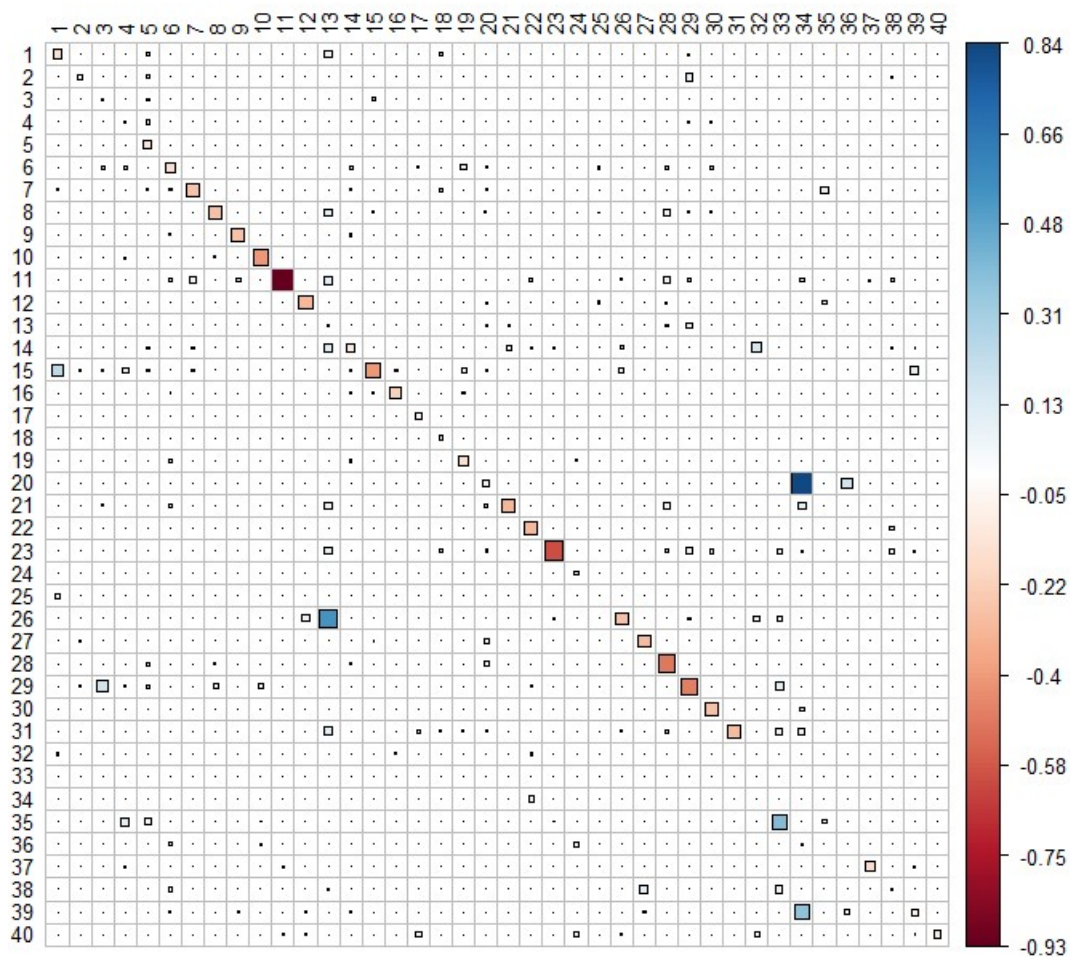


Fig.2. The average own-price elasticities and cross-price elasticities across 10 rolling periods

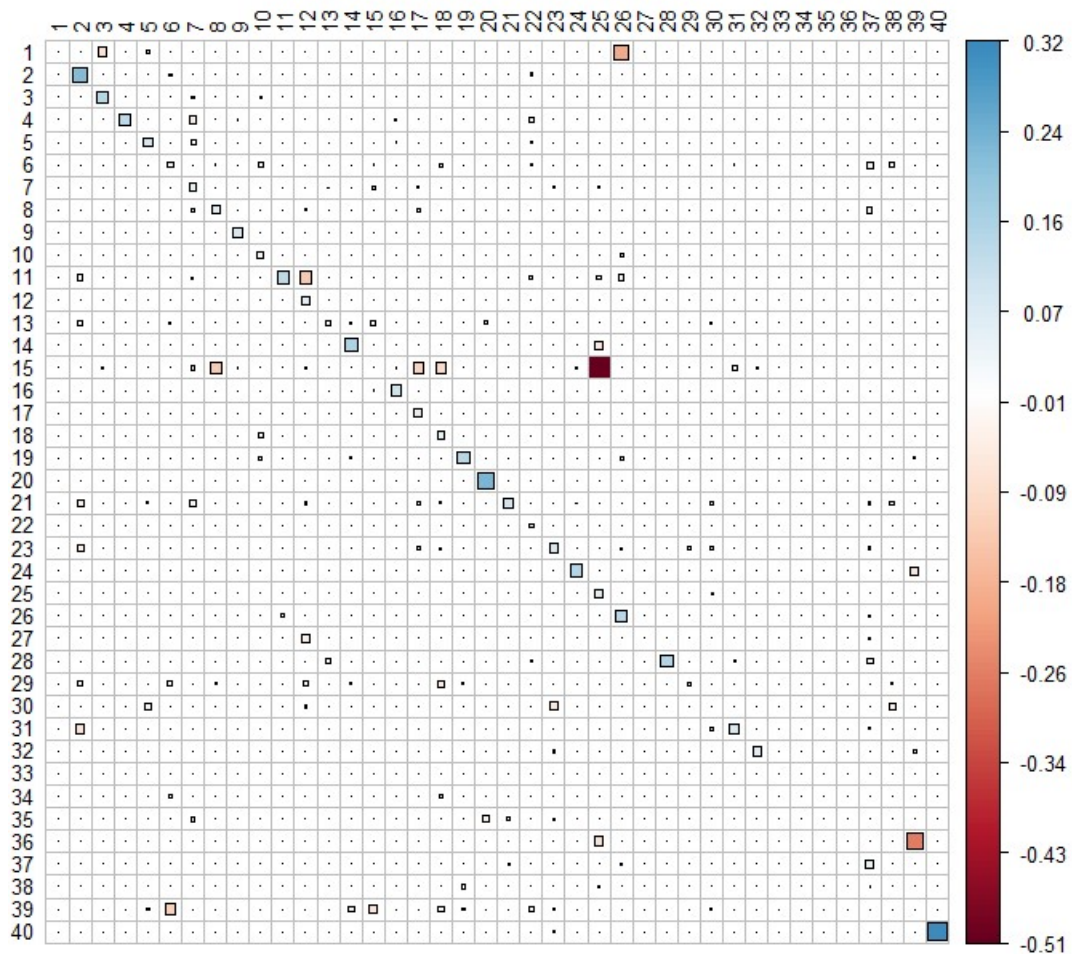


Fig.3. The average own-effects and cross-effects of feature advertising

Considering the large scale of total number of parameters we generated, we cannot report all the estimated parameters in detail. In Fig. 2, we depict the average own-price and cross-price elasticities across 10 rolling periods for the first category in store 4. These elasticities equal the negative of the coefficients in the model divided by average prices. In Fig.3, we depict the average own feature advertising effects and cross feature advertising effects. In Fig.2 and 3, both the horizontal and vertical labels are the ID of SKUs in the category (i.e., 40 SKUs in category 1, store 4). So the diagonal cells represent the own elasticities, and the others represent cross elasticities. The color legends in the figures show the direction and magnitude of the elasticity. Blue represents positive and red means negative, and the deeper the color means the bigger elasticity. Based on the Fig.2 and 3, store managers could empirically understand how the promotions on one SKU influence the sales of its own and others, which could be helpful in planning the weekly promotion schedule. For instance, from Fig.2 and 3, we learn that the price reduction of SKU-34 would decrease the sales of SKU-20, but no evidence in the data to show that the feature advertising of SKU-34 affects

the sales of SKU-20.

The results show there is considerable heterogeneity for own-price elasticity among different SKUs, ranging from -2.9 to 0: also that the cross-price elasticity between pairs of SKUs is asymmetric. For example, SKU-34 is affected by the price reduction of SKU-20, but not vice versa. Fig.2 also shows the sparsity in the cross-price elastics among SKUs. On an average, an SKU's price change may impact on the sales of just five other SKUs in the category.

We use Mean Absolute Error (MAE), Mean Absolute Scaled Error (MASE) and Mean Percentage Error (MPE) to compare the forecasting performance of the candidate demand models. MAE is a traditional and popular scale-dependent error measures which is easy to calculate, easy to understand and widely applied. The Mean Absolute Scaled Error (MASE) which was proposed by Hyndman and Koehler (2006), can be considered as a “weighted” arithmetic mean of the MAE based on the variations of the sales data in the estimation period (Davydenko and Fildes, 2013). MPE is used to measure the forecasting error bias, potentially important in stocking decisions, and is defined here as the arithmetic mean of ratios of total error to total sales per SKU.

In Table 4, we compare the forecasting accuracy for the various demand models introduced in section 3. In general, the ADL-CROSS model outperforms ADL-OWN as well as the two other alternative models across all the performance measures. The Promotional index model also performs pretty well: though it is not the best on average, it provides most accurate forecasts in 4 store categories, i.e. beer in stores 1 and 3, mayo in stores 2 and 4. The TOP-DOWN decomposition model generates the worst forecasts for most store categories, except for Mayo in store 3. For none of categories, does the ADL-OWN model provide the best forecasts. These results show that ADL-CROSS is superior over all other alternative models and integrating cross SKU promotional competitive information is potentially important in improving SKU demand forecasting accuracy. Overall, however, its forecasts are negatively biased.

Table 4

Forecasting validation for various demand models

Store	Category	ADL-OWN model			ADL-CROSS model			TOP-DOWN model			Promotional index model		
		MAE	MASE	MPE	MAE	MASE	MPE	MAE	MASE	MPE	MAE	MASE	MPE
1	Milk	9.720	1.018	-1.29	9.228	0.975	-1.08	10.531	1.052	0.12	9.415	0.997	-1.72
	Beer	2.209	0.811	-7.40	2.062	0.756	-1.72	2.053	0.753	0.01	2.067	0.756	-2.42
	Mayo	2.483	0.896	-7.49	2.349	0.842	0.47	2.281	0.834	-0.49	2.406	0.862	-2.60
	Yogurt	9.532	0.862	-4.81	8.943	0.802	-3.01	9.466	0.826	-4.23	9.186	0.826	-4.01
2	Milk	15.943	0.938	-2.26	15.682	0.914	-2.44	20.258	1.137	-1.56	15.901	0.933	-2.62
	Beer	3.285	0.814	-2.12	3.122	0.776	0.63	4.037	1.058	-6.52	3.150	0.779	-0.82
	Mayo	8.424	1.331	-20.99	6.342	0.785	-7.39	6.548	0.798	-2.35	6.237	0.800	-8.53
	Yogurt	22.154	0.696	-3.58	21.121	0.666	-0.25	22.768	0.744	1.05	22.033	0.685	-2.65
3	Milk	24.593	0.898	-4.50	24.186	0.868	-3.04	30.934	1.076	-4.80	25.318	0.920	-5.19
	Beer	2.516	0.765	-4.61	2.456	0.733	-1.36	2.674	0.791	-3.01	2.419	0.724	-2.37
	Mayo	10.453	0.920	-6.27	9.696	0.885	-2.74	8.859	0.901	-3.24	9.557	0.896	-5.59
	Yogurt	26.015	0.649	-5.97	25.179	0.633	-4.54	27.473	0.702	-3.32	25.318	0.920	-5.22
4	Milk	7.489	0.844	-3.26	7.325	0.809	-2.40	10.702	1.073	3.45	7.857	0.848	-0.94
	Beer	4.246	0.868	0.68	4.092	0.830	2.79	4.607	0.916	1.07	4.133	0.838	4.38
	Mayo	4.196	0.765	-3.33	3.976	0.722	-2.39	4.059	0.749	-3.09	3.946	0.715	-3.44
	Yogurt	14.893	0.757	-3.11	14.638	0.735	-1.25	14.992	0.791	-2.95	14.885	0.751	-2.76
Average		10.509	0.865	-5.02	10.025	0.796	-1.86	11.390	0.888	-1.87	10.239	0.828	-2.91

5.3 Promotion optimization experiments

We first conduct a promotion optimization experiment on all 4 categories in 4 stores in a typical set of cost settings. In order to closely imitate a realistic application, we first estimate the ADL-CROSS model with the first 200 weeks data and then optimize the promotional schedule for the following 8 weeks planning horizon. These optimal promotional schedules are then used to produce an estimate of the sales and profits over the horizon. The dataset does not provide information on SKUs' unit cost, label changing cost, nor the cost of display or feature advertising. In the following simulation optimization experiments, the gross margin is assumed as 20% of the price for each SKU over the planning horizon. We also arbitrarily set the cost of display as \$20, the feature advertising as \$20 per SKU per week, and the price retagging cost as \$1. To avoid the extreme prices in the solution, the feasible price set for each SKU is obtained by extracting the unique prices at quantiles of 20%, 35%, 50%, 65% and 80% from each SKU's price history in the planning horizon. To keep the stores' price image unchanged, we set the lower bound of the allowed number of SKUs in price discounts in the category, i.e., Lp_t , as their historical mean of the category in the data (fifth column in the Table 3). The upper bounds for the number of SKUs discounted in a category, as well as the display and feature advertising in a given week in the category is set at their historical maximum value of the category in the data. Similarly, for each SKU, we assume the limitations for the frequency of price reduction, display and feature advertising across planning horizon are also the same as their historical maximum value in the data. For example, the maximal frequency of the display over the planning horizon is $Max(\sum_{t=1}^T d_{jt} / T), \forall j$ in the data. We used such bounds in our simulative experiments to make a fair comparison between optimal promotion planning and real ones. These values are summarized in Table 5. The lower bound for display and features are all set as zero.

The genetic algorithm proceeds by searching for the optimal level of price, display and feature advertising for each SKU that maximizes category profit in planning horizon. We set the parameters in the genetic algorithm as follows: population size = 1000, probability of crossover = .8, probability of mutation = .25, the number of best fitness individuals to survive at each generation is set as 5% and convergence criteria = 100 consecutive generations without any improvement in the best fitness value.

The comparative results from the actual and optimal promotion schedule are presented in Table 6. Under the experimental settings, when adopting the optimal schedule, we expect to obtain \$26.91 profit from each SKU per week on average, while actual value is \$22.75. This means that the optimal schedule is expected to increase profit by 18%, or \$33,452 for a total of 915 SKUs over an 8 weeks planning period. The predicted profit (and sales) in Table 6 are generated by ADL-CROSS model using actual promotion schedule as the inputs. Both predictions of profit and sales for all the categories are very close to the actual values. On average, the optimal sales are expected to be a little higher than the actual sales, but the price is decreased by \$0.06 on average, and less frequent displays and feature advertisings are recommended.

Table 5
The upper bounds of promotions in the optimization experiments

Store	category	Max Promotion Frequency over the planning horizon			Max Promotion Density in each week		
		Discounts	Displays	Features	Discounts	Displays	Features
1	Milk	0.42	0	0.57	0.64	0	0.61
	Beer	0.59	0.19	0.33	0.57	0.17	0.25
	Mayo	0.54	0.3	0.25	0.7	0.2	0.7
	Yogurt	0.74	0	0.29	0.82	0	0.64
2	Milk	0.23	0.01	0.21	0.49	0.03	0.25
	Beer	0.43	0.31	0.45	0.17	0.1	0.14
	Mayo	0.25	0.11	0.18	0.42	0.16	0.26
	Yogurt	0.54	0.03	0.18	0.57	0.14	0.44
3	Milk	0.32	0.14	0.16	0.55	0.11	0.4
	Beer	0.44	0.12	0.13	0.5	0.4	0.4
	Mayo	0.34	0.49	0.1	0.43	0.19	0.14
	Yogurt	0.59	0.18	0.16	0.96	0.33	0.54
4	Milk	0.39	0.01	0.17	0.6	0.08	0.68
	Beer	0.75	0.85	0.51	0.49	0.44	0.32
	Mayo	0.51	0.18	0.26	0.59	0.24	0.76
	Yogurt	0.78	0.04	0.13	0.98	0.19	0.96

Table 6
Results of simulative promotion optimization experiments

Store	Category	Avg. profit			Avg. price		Avg. features & displays		Avg. sales		
		Optimized	Predicted	Actual	Optimized	Actual	Optimized	Actual	Optimized	Predicted	Actual
1	Milk	60.47	51.70	52.59	2.76	2.80	0.00	0.14	116.20	109.86	111.49
	Beer	7.90	4.47	3.42	7.83	7.96	0.02	0.10	5.38	4.72	4.28
	Mayo	2.46	1.44	1.51	3.59	3.23	0.00	0.01	4.58	5.03	5.03
	Yogurt	6.10	2.10	3.16	1.53	1.72	0.00	0.18	33.56	31.42	38.87
2	Milk	83.74	72.68	76.27	2.63	2.52	0.03	0.05	154.24	150.54	157.77
	Beer	17.24	14.72	14.55	8.19	8.24	0.06	0.07	9.66	8.45	8.61
	Mayo	9.57	8.00	8.17	2.66	2.64	0.01	0.03	19.69	19.10	19.82
	Yogurt	12.70	9.70	11.28	1.26	1.16	0.02	0.13	75.06	83.02	92.11
3	Milk	132.58	118.87	114.26	3.05	2.99	0.05	0.03	201.73	185.13	176.13
	Beer	6.53	6.47	5.72	6.70	6.78	0.00	0.03	5.33	5.48	4.84
	Mayo	16.88	14.43	13.05	2.90	2.76	0.07	0.01	34.53	31.44	28.92
	Yogurt	17.44	14.40	14.63	1.41	1.35	0.05	0.08	109.10	110.00	110.62
4	Milk	23.40	22.06	23.29	2.81	2.88	0.02	0.04	53.65	47.98	50.28
	Beer	18.77	6.94	7.88	7.64	8.14	0.02	0.14	12.42	10.34	11.75
	Mayo	6.54	5.12	5.98	2.80	3.14	0.00	0.04	14.60	11.83	13.38
	Yogurt	8.20	7.94	8.21	1.33	1.62	0.00	0.04	43.75	40.99	45.52
Average		26.91	22.56	22.75	3.69	3.75	0.02	0.07	55.84	53.46	54.96

No general conclusions as to the optimal pricing/ promotions strategies are possible as they depend critically not just on the relative elasticities but also on the marginal costs and profitability. To understand the impact of the cost settings on the optimal profit lifts over the actual value, we study the effect of each of these factors by varying them one at a time while the others are set to the same value as in the previous section. Specifically, in Fig.4, we vary the display and feature advertising cost, i.e., C_{j2} and C_{j3} in (8), from \$0 to \$30 per SKU per week, the price margin $(1-C_{j1}/p_{jt})$ from 5 percent to 35 percent, and the price tag changing cost C_{j4} from \$ 1 to \$3 per SKU per changing. Fig.4 leads to the following observations: 1) the improvements of the optimal profit over the actual profit are bigger when the display and feature advertising costs are higher; 2) the lifts of the optimal profit over the actual profit are smaller when the profit margin and label costs are higher; 3) the optimal profits are consistently higher over the actual profit in all the cost settings: the minimum profit lift is 13.6%. We also test the worst situation by setting the display and feature advertising cost as \$0, and setting the profit margin and retagging costs as 35% and \$3 respectively: and we still get 10.6% profit lifts from the proposed optimization algorithm.

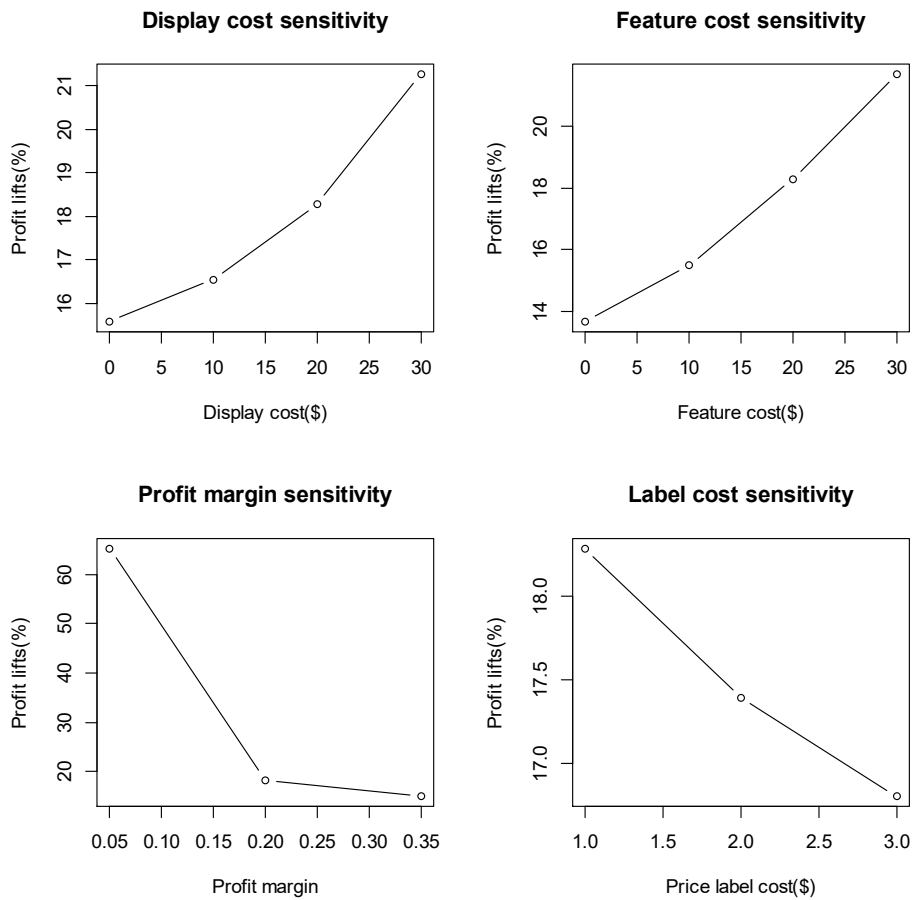


Fig. 4. Sensitivity analyses of cost factors on profit lifts

In order to check robustness, we have tested the approach by estimating the model and optimizing the promotional schedule in a rolling fashion. Specifically, we use 200 weeks as a rolling window to estimate the model parameters, and then optimize the following 8 weeks' promotional schedules. The rolling window is kept updated and moved forward for every 8 weeks. This rolling process is implemented for five times to simulate a realistic promotion schedule process for the last 40 weeks of the data. All the costs and boundaries needed in the model are set at the same level as in the first optimization experiment. Fig.5 reports the results. For all the five planning horizons, we get similar results to that for the first optimization planning horizon. The optimal profits are consistently larger than the actual profits. The profit lifts are in between 12% to 21% and the average profit lifts is 16.7%.

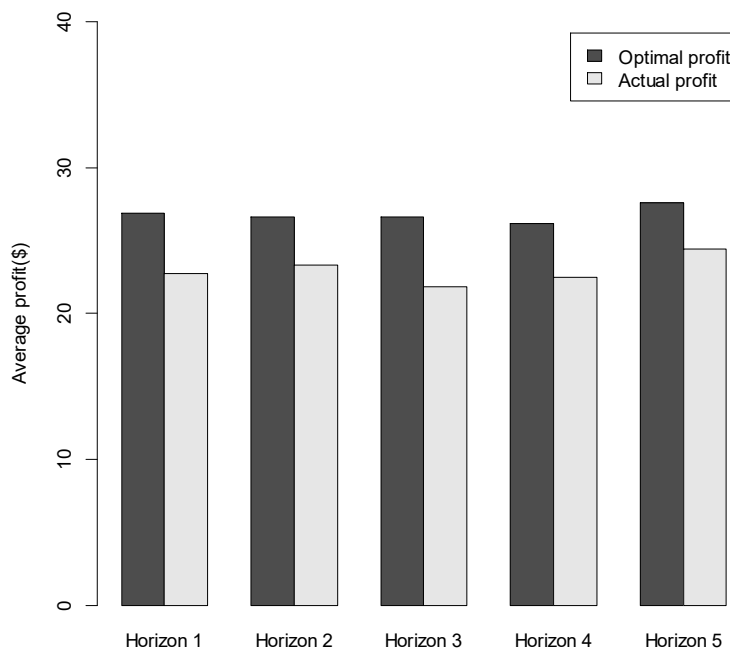


Fig.5. The robustness checks with rolling planning optimization

Another interesting question to explore is the benefits of including cross SKU promotional effects, compared to only considering own promotional effects: how much profits would possibly be lost? To simulate this situation, we repeat the first promotions optimization experiments as described, but using the ADL-OWN model and its parameters as the demand model in the optimization instead of ADL-CROSS. After obtaining the optimal promotion schedule, we input the schedule into the ADL-CROSS again, using ADL-CROSS to generate predicted profits which are reported in Table 7. Results show that, by neglecting

cross SKU promotional effects, we expect to lose the profit \$5451 on 915 SKUs over 8 weeks planning horizon, or 16.3 percent of the total profit lifts expected to be obtained. Similarly, there is a question as to the benefits of considering cross-period effects. For each SKU in the sample, we estimate a Log Linear Model which is similar with the ADL-CROSS but exclude all its dynamic components (promotion lags, $\ln(\tilde{Y}_{kjt})$ and $\ln(Y_{ki,t-1})$) in Eq.(1). After obtaining the optimal promotion schedule by using the Log Linear Model as the demand model, we input the schedule into the ADL-CROSS again, using ADL-CROSS to generate predicted profits which are also reported in Table 7. By neglecting cross period promotional effects, we expect to lose profits of \$5237 in total, about 15.6 percent of the total profit lifts obtained. These indicate that integrating both cross-item effects and cross-period effects are important in a promotions optimization DSS.

Table 7
The value of the cross-item and cross-period promotional effects

Store	Category	Profit lose by neglecting cross-item effects (\$)	Profit lose by neglecting cross-period effects (\$)
1	Milk	258.8	231.5
	Beer	458.7	159.0
	Mayo	4.8	8.0
	Yogurt	34.4	11.1
2	Milk	192.8	377.3
	Beer	1724.7	2255.7
	Mayo	61.8	43.1
	Yogurt	219.6	400.7
3	Milk	616.6	750.6
	Beer	4.0	10.2
	Mayo	230.0	127.1
	Yogurt	411.1	438.1
4	Milk	150.4	123.7
	Beer	849.3	242.4
	Mayo	41.5	9.2
	Yogurt	192.7	49.1
Total		5451.0	5236.7

6. Discussion and conclusion

We have proposed a dynamic promotions planning model at SKU level that gives the optimal timing of promotions and depth of retail discounts under the constraints arising from imposing important business rules. Those decisions include when, for how long, and, in the case of temporary price reductions, how deep to run promotional events. We link the demand model that captures both the promotional cross-product effects and cross-period influences to an optimization module that uses the Genetic algorithm to search for the set of decisions over the planning horizon that maximizes category profit. A unique aspect of our framework is that it considers all the basic requirements for a promotion DSS.

As we have argued, this is the first paper which aims at retail store long term category level profit maximization by collaboratively promotions planning at SKU level. We provide retailers with a practicable SKU level promotion DSS tool which is capable of providing accurate sales forecasting and optimal promotions schedule simultaneously. Our substantive applications demonstrate that by integrating both cross-product and cross-period promotion influences we can significantly improve forecasting accuracy over the model that considers only cross-period influences. The optimal promotion schedules are expected to increase profits by 16.7% over the actual implemented on average. We can obtain on average 16.3 % more profit over the model that considers only own promotional influences, or 15.6% over the model that neglects promotional cross period influences. The results have proved to be robust across various cost parameterizations. From these we can conclude that both cross-item and cross-period promotional information are valuable in SKU level sales forecasting and promotions planning, and retailers could achieve greater profit by aiming at whole category multi-periods profit maximization instead of at individual item short period gains.

The research has certain limitations which would benefit from further analysis. While the data set is well established, extending the modelling to more stores and other data sets of different types (e.g. D-I-Y to compare with Natter et al. 2007) would help practitioners resolve questions of relevance to their own problem areas. Further research questions include extending our general model to include other realistic aspects that may reflect actual retailer discounting practices. For example, by considering order and stock cost. We can also fix some of SKU's price while optimizing others if the deal is launched by a manufacturer.

Another extension of our model is to integrate cross-category promotions effects to optimize store level profit. The practical limitations of our overall model are few. The optimization model is flexible enough to be applicable in any situation, as long as valid demand forecasts can be obtained. The demand models proposed in this paper require enough historical data and relatively stable assortment. For new SKUs, or SKUs with short sales history, the proposed demand models will not be immediately applicable. Some variations of the model are needed for dealing with these situations. For example, the demand of new SKU could be evaluated by reference to analogous SKUs with similar attributes (flavor, package, color, ingredients, etc.). There are also computational issues where the intensive use of the genetic algorithm to solve the optimization problem in practical applications is overly time consuming – the replacement heuristic algorithm could be further developed to speed up the optimization process. Finally, there is the fascinating problem of introducing just such a DSS into a retail chain and changing its promotional practices.

Acknowledgments The first author acknowledges the ongoing support of the National Natural Science Foundation of China under grant nos. 71171100, 71571089.

References

- Agrawal, N., & Smith, S. A. (2013). Optimal inventory management for a retail chain with diverse store demands. *European Journal of Operational Research*, 225(3), 393-403.
- Ailawadi, K. L., Harlam, B. A., César, J., & Trounce, D. (2007a). Practice Prize Report—Quantifying and Improving Promotion Effectiveness at CVS. *Marketing Science*, 26(4), 566-575.
- Ailawadi, K. L., Gedenk, K., Lutzky, C., & Neslin, S. A. (2007b). Decomposition of the sales impact of promotion-induced stockpiling. *Journal of Marketing Research*, 44(3), 450-467.
- Blattberg, R.C., & Neslin S. C. (1990). Sales promotions: concepts, methods, and strategies. Englewood Cliffs, NJ: PrenticeHall.
- Bolton, R.N., Shankar V. & Montoya D. (2007). Recent trends and emerging practices in retail pricing, in retailing in the 21st century: current and future trends, Kraft M. and Mantrala M., eds (second edition). Germany: METRO
- Bronnenberg B.J., Kruger, M.W., & Carl, F. M. (2008). The IRI Academic Dataset, *Marketing Science*, 27(4), 745-748.
- Bucklin, R.E., Gupta, S.(1999). Commercial use of UPC scanner data: industry and academic perspectives. *Marketing Science*, 18(3) 247-273.
- Chapados, N., Joliveau, M., L'Ecuyer, P., & Rousseau, L.M. (2014). Retail store scheduling for profit. *European Journal of Operational Research*, 239(3), 609-624.
- Cohen, M.C., Leung, Z., Panchangam, K., Perakis, G., & Smith, A. (2014). The impact of linear optimization on promotion planning, MIT working paper.

- Cooper, L. G. (1993). Market-Share Models. Handbook in Operations Research and Management Science, North-Holland, Amsterdam.
- Cooper, L. G. & Nakanishi M. (1988). Market share analysis: evaluating competitive marketing effectiveness, Kluwer Academic Publishers, Boston.
- Davydenko, A., & Fildes, R. (2013). Measuring forecasting accuracy: The case of judgmental adjustments to SKU-level demand forecasts. *International Journal of Forecasting*, 29(3), 510–522.
- Dorsey, R. E. & Mayer W.J. (1995). Genetic algorithms for estimation problems with multiple optima. *Journal of Business and Economic Statistics*, 13 (1), 53–67.
- Fan J, Lv J. (2008). Sure independence screening for ultrahigh dimensional feature space. *Journal of Royal Statistical Society, Series B*, 70, 849–911.
- Fok, D., Franses P. H., & Paap R. (2003). Modeling dynamic effects of the marketing mix on market shares, ERIM Report Series in Management ERS-2003-44-MKT, Erasmus University, Rotterdam.
- Friedman, J., Hastie, T., & Tibshirani, R. (2010). Regularization Paths for Generalized Linear Models via Coordinate Descent, *Journal of Statistical Software*, 33(1), 1-22.
- Gedenk, K., Sattler H. (1999). The Impact of Price Thresholds on Profit Contribution. Should Retailers Set 9-Ending Prices? *Journal of Retailing*, 75(1), 33-57.
- Goldberg, D. E. (1989). A Simple Genetic Algorithm. In Genetic algorithms in search optimization and machine learning . Boston: Addison-Wesley Longman, 10–14.
- Guadagni, P.M., Little J.D.C. (1983). A logit model of brand choice calibrated on scanner data. *Marketing Science*, 2(summer), 203-238.
- Gupta, S. (1988). Impact of sales promotion on when, what and how much to buy. *Journal of Marketing Research*, 25 (November), 342-355.
- Gür Ali, Ö., & Yaman, K. (2013). Selecting rows and columns for training support vector regression models with large retail datasets. *European Journal of Operational Research*, 226(3), 471-480.
- Hyndman, R. J., & Koehler, A. B. (2006). Another look at measures of forecast accuracy. *International Journal of Forecasting*, 22, 679-688.
- Huang, T., Fildes, R. & Soopramanien, D. (2014). The value of competitive information in forecasting FMCG retail product sales and the variable selection problem. *European Journal of Operational Research*, 237(2), 738-748.
- Hübner, A., Kuhn, H., & Kühn, S. (2016). An efficient algorithm for capacitated assortment planning with stochastic demand and substitution. *European Journal of Operational Research*, 250(2), 505-520.
- Ferreira, K. J., Lee, B. H. A., & Simchi-Levi, D. (2015). Analytics for an online retailer: demand forecasting and price optimization. *Manufacturing & Service Operations Management*, 18(1), 69-88.
- Kunz, T. P. ; Crone, S.F (2014). Demand models for the static retail price optimization problem - A Revenue Management perspective. 4th Student Conference on Operational Research, Nottingham, UK.
- Kunz, T. P., & Crone, S. F. (2015). The impact of practitioner business rules on the optimality of a static retail revenue management system. *Journal of Revenue & Pricing Management*, 14(3), 198-210.
- Lang, S., Steiner, W. J., Weber, A., & Wechselberger, P. (2015). Accommodating heterogeneity and nonlinearity in price effects for predicting brand sales and profits. *European Journal of Operational Research*, 246(1), 232-241.
- Levy, M., Grewal D., Kopalle P. K., Hess J. D. (2004). Emerging trends in retail pricing practice: Implications for research. *Journal of Retailing*, 80(3), xiii-xxi.
- Liang, F. & Wong, W.H. (2001). Real-parameter evolutionary Monte Carlo with application to Bayesian mixture models. *Journal of the American Statistical Association* , 96 (454), 653–66.

- Lan W., Yang Y., Liu H. (2014), Nonnegative-lasso and application in index tracking. *Computational Statistics & Data Analysis*, 70(2), 116-126.
- Ma, S., Fildes, R., & Huang, T. (2016). Demand forecasting with high dimensional data: The case of SKU retail sales forecasting with intra- and inter-category promotional information. *European Journal of Operational Research*, 249(1), 245-257.
- Macé, S. and Neslin S.A (2004). The determinants of pre-and postpromotion dips in sales of frequently purchased goods. *Journal of Marketing Research*, 41(3), 339-350.
- Mela, C.F., Kamel J., & Douglas B. (1998). The long-term impact of promotions on consumer stockpiling behavior. *Journal of Marketing Research*, 34 (2), 250-262.
- Montgomery, A. L. (2005). The implementation challenge of pricing decision support systems for retail managers. *Applied Stochastic Models in Business and Industry*, 21(4-5), 367-378.
- Mullet, G. (1976). Why regression coefficients have the wrong sign. *Journal of Quality Technology*. 8(3), 121-126.
- Mulhern, F. J., & Leone, R.P. (1991). Implicit price bundling of retail products: A multiproduct approach to maximizing store profitability. *Journal of Marketing*, 55(October), 63–76.
- Natter, M., Reutterer, T., Mild, A., Taudes, A. (2007) Practice prize report-an assortmentwide Decision-Support System for dynamic pricing and promotion planning in DIY retailing. *Marketing Science*, 26(4), 576-583
- Nissen, V. (1995) An overview of evolutionary algorithms in management applications. In: Biethahn, J.; Nissen, V. (eds.) *Evolutionary algorithms in management applications*, Springer, 44 – 97.
- Raju, J. S. (1995). Theoretical models of sales promotions: Contributions, limitations, and a future research agenda. *European Journal of Operational Research*, 85(1), 1-17.
- Reibstein, D.J. and Gatignon, H. (1984). Optimal product line pricing: the influence of elasticities and cross-elasticities. *Journal of Marketing Research*, 21(August): 259-267.
- Rinne, H., & Geurts, M. (1988). A forecasting model to evaluate the profitability of price promotions. *European Journal of Operational Research*, 33, 279-289.
- Scrucca, L. (2013). GA: A Package for Genetic Algorithms in R. *Journal of Statistical Software*, 53(4), 1-37
- Silva-Risso, J.M., Bucklin R. & Morrison D. (1999). A Decision Support System for planning manufacturers' sales promotion calendars, *Marketing Science*, 18 (3), 274-300.
- Sun, B. (2005). Promotion effects on endogenous consumption. *Marketing Science*, 24 (3), 430-443.
- Sun, B., Neslin S.A. & Kannan S. (2003). Measuring the impact of promotions on brand switching when consumers are forward looking. *Journal of Marketing Research*, 40 (4), 389-405.
- Tellis, G.J., & Zufryden, F. (1995). Tackling the retailer decision maze: Which brand to discount, how much, when and why? *Marketing Science*, 14(3), 271–299.
- Tibshirani, R. (1996). Regression shrinkage and selection via the LASSO. *Journal of Royal Statistical Society, Series B*, 58(1), 267-288.
- Tibshirani, R. (2011). Regression shrinkage and selection via the lasso: a retrospective. *Journal of the Royal Statistical Society: Series B*, 73(3), 273-282.
- Van Heerde H., Gupta S., & Wittink D.R. (2003). Is 75% of the sales promotion bump due to brand switching? No, only 33% is. *Journal of Marketing Research*, 40(4), 481-491.
- Van Heerde H., Leeflang P. S.H., & Wittink D.R. (2000). The estimation of pre- and postpromotion dips with store-level scanner data. *Journal of Marketing Research*. 37(3), 383-395.
- Van Heerde H., Leeflang P. S.H., & Wittink D.R. (2002). How promotions work: SCAN*PRO-based evolutionary model building. *Schmalenbach Business Review*, 54(3), 198–220.
- Van Heerde H., Leeflang P. S.H., & Wittink D.R. (2004). Decomposing the sales promotion bump with store

- data. *Marketing Science*, 23(3), 317-334.
- Venkatesan, R., Krishnan, T.V., & Kumar V. (2004). Evolutionary estimation of macro-level diffusion models. *Marketing Science*, 23(3):451-464
- Venkatesan, R.& Kumar, V. A customer lifetime value framework for customer selection and optimal resource allocation strategy. *Journal of Marketing*, 2004, 68(4):106-125.
- Vilcassim, N.& Chintagunta,P. (1995). Investigating retailer product category pricing from household scanner panel data. *Journal of Retailing*, 71 (2), 103-128.
- Williams, B. D., Waller, M. A., Ahire, S., & Ferrier, G. D. (2014). Predicting retailer orders with POS and order data: The inventory balance effect. *European Journal of Operational Research*, 232(3), 593-600.
- Zellner, A. (1962), An efficient method of estimating seemingly unrelated regressions and tests of aggregation bias. *Journal of the American Statistical Association*, 57, 348–368.