

Interaction Between Shelf Layout and Marketing Effectiveness and its Impact on Optimizing Shelf Arrangements

Erjen van Nierop

Faculty of Economics and Business, Department of Marketing, University of Groningen,
9700 AV Groningen, The Netherlands, j.e.m.van.nierop@rug.nl

Dennis Fok, Philip Hans Franses

Econometric Institute, Erasmus University Rotterdam 3062 PA Rotterdam, The Netherlands
{dfok@few.eur.nl, franses@few.eur.nl}

In this paper, we propose and operationalize a new method for optimizing shelf arrangements. We show that there are important dependencies between the layout of the shelf and stock-keeping unit (SKU) sales and marketing effectiveness. The importance of these dependencies is further shown by the substantive profit gains we obtain with our proposed shelf optimization approach. The basis of our model is a standard sales equation that explains sales using item-specific marketing effect parameters and intercepts. In a Hierarchical Bayes (HB) fashion, we augment this model with a second layer that relates the effect parameters to shelf and SKU descriptors. We also take into account potential endogeneity of facings. After estimating the parameters of the two-level model using Bayesian methodology, we carefully investigate the dependencies of SKU sales and SKU marketing effectiveness on the shelf layout. Next, we search for the shelf arrangement that maximizes the expected total profit using simulated annealing (SA). We appear to be able to increase profits for all the stores analyzed, and our approach appears to outperform well-known rules of thumb.

Key words: shelf management; sales models; Hierarchical Bayes; Markov Chain Monte Carlo; simulated annealing

History: This paper was received March 26, 2006, and was with the authors 10 months for 2 revisions; processed by Eric Bradlow. Published online in *Articles in Advance*.

1. Introduction

Retailers have limited shelf space available. The choice of which items to stock and the allocation of scarce shelf space among the stocked items are relevant issues for the retailer. For individual SKUs these decisions are important determinants of sales and marketing effectiveness. At the aggregate level, shelf allocation is an important factor in the revenue, cost, and eventual profit of a product category.¹ Complementary to the amount of space to allocate to an item, there is the problem of the location of the item on the shelf. For example, items on the lower shelf usually get less consumer attention than items on upper shelves. The items on the lower shelves may therefore have lower sales and may also benefit less from promotions.

Finding the profit-maximizing shelf arrangement while, at the same time, meeting manufacturers

requirements is far from easy. A prerequisite to actual shelf optimization is a proper measurement of the effect of shelf layout on sales and marketing effectiveness. An adequate shelf management model would be a very useful aid to retailers to estimate these relations and to support their decisions and negotiations with manufacturers.

In this paper, we propose such a shelf management model. Our approach consists of two parts. In the first, we provide a statistical model to measure the direct and indirect impact of shelf layout on sales. The second part is a tool to search for the profit-maximizing shelf layout. The basis of our statistical model is a standard sales equation that explains (the logarithm of) sales from item-specific marketing effect parameters and intercepts. In an HB fashion, this model is augmented with a second equation that relates the marketing effect parameters to shelf and SKU descriptors. This second equation provides the link between shelf allocation on the one hand and sales and marketing effectiveness on the other. Additionally, we take into account the endogeneity of the number of facings. We estimate the parameters of

¹ The scarcity of shelf space also influences the distribution of power in the retail chain (see, for example, Luo et al. 2007 and Geylani et al. 2007).

the two-level model using the Bayesian methodology, in particular Gibbs sampling. The estimated model parameters measure the effect of shelf layout on baseline sales and on the effectiveness of marketing instruments such as price and promotions. We use graphs to visualize these (nonlinear) effects. To investigate how the model performs in forecasting sales, we predict sales for a hold-out sample of five weeks of data. Furthermore, and most important, the model is used to optimize shelf allocation. For this, we consider SA, for its ease of implementation and the ability to search across a large and complex solution space as well as for its ability to avoid getting stuck in a local optimum.

The remainder of this paper is organized as follows. In §2, we review the current literature on shelf management. We also indicate the added value of our approach with respect to the current literature. In §3, we discuss our approach. Next, we present the technicalities of our model in §4. Section 5 illustrates our shelf management approach using a database of the canned soup category. This section includes a rich description of the shelf space and location on the shelf of a large number of products, where shelf layouts were manipulated in an experimental setting. We conclude in §6.

2. Literature

In the 1960s and 1970s, a number of experiments were conducted to measure the effect of shelf space on sales, (see, for example, Brown and Tucker 1961, Cox 1970, and Curhan 1972). These authors only considered the problem of measuring this effect. Models to (partly) solve the shelf management problem have been proposed in past decades. Corstjens and Doyle (1981, 1983) were the first to optimize store profitability with respect to space allocation. They consider both the main and the cross-space elasticities in their multiplicative demand function, and specify a cost function that moderates the profitability of the allocation. This shelf-space optimization problem is then solved within a geometrical programming framework. In a comparison of their approach with alternative procedures, they find that their general model leads to significantly different allocation rules and better profit performance.

Bultez and Naert (1988) build on the work of Corstjens and Doyle (1981, 1983) in their Shelf Allocation for Retailer's Profit (SH.A.R.P.) model. The authors derive an expression for the optimal shelf space to be allocated to an SKU. This expression depends on the cross-space elasticities between the items. Commonly used rules of thumb for space allocation are compared and shown to be special, though inferior, cases of the optimal rule are derived. The

authors apply the model to experimental data with six brands and find that the proposed model improves on current profit levels and that it is better than the rules of thumb. However, the optimization focuses only on the shelf space devoted to an item, and does not include other shelf layout descriptors such as shelf height and the horizontal position of an item on the shelf, nor does it include marketing instruments such as feature and price.

Drèze et al. (1994) conduct a series of field experiments in which they measure the effectiveness of two shelf management techniques: *space to movement*, where the shelf is customized based on historic store-specific movement patterns, and *product reorganization*, where product placement is manipulated to facilitate cross-category merchandizing or ease of shopping. The authors find sales gains of about 4% with the first manipulation and 5% to 6% with the second. The impact of shelf positioning and facing allocations on sales of individual items is also analyzed. In particular, location appears to have a large impact on sales. For example, in most categories, products perform best when placed at eye level.

Borin et al. (1994) develop a category management model formulated as a constrained optimization problem, with assortment and allocation of space as the decision variables. The model parameters are based on judgmental estimates, that is, they are not based on an econometric model. In the next step, the authors use SA to improve the shelf layout for two data sets. The two data sets analyzed contain 6 SKUs and 18 SKUs, respectively. In a followup study (Borin and Farris 1995), the authors examine the sensitivity of the analysis to errors in the judgements. More specifically, they find the maximum degree of error that may be introduced before the model yields assortments and shelf allocation that are inferior compared to those produced by the merchandizing rule of thumb to set share-of-shelf equal to share of sales. Their results show that as much as 50% variation in the estimates of parameters is allowed before the model appears unusable.

In a more or less separate stream of research, optimization routines for shelf allocation have been investigated. Although several routines have been proposed to optimize shelf layouts, they often only consider the shelf number (vertical position) and the number of facings to allocate. Furthermore, the indirect effect of the layout on sales through marketing effect parameters is not considered. Yang and Chen (1999) use a simplified version of the integer programming model of Corstjens and Doyle (1981), whereas Yang (2001) uses the knapsack algorithm. Lim et al. (2004) build on this work by optimizing profits with two metaheuristic approaches: Tabu Search and the so-called Squeaky Wheel Optimization. Although

their method appears to outperform Yang’s (2001) heuristic, by using simulated sales data and fixed and known parameters, these approaches assume that the effect of shelf layout on sales is given. In a real-life situation, this is of course not true and one needs to estimate the relation between sales and shelf layout for a particular situation.

In this paper, we propose an HB model to estimate the interaction between shelf layout and sales and between shelf layout and marketing instrument effectiveness. In an HB model, the parameters for individual items are assumed to be samples from a common distribution, where the mean of the distribution depends on the particular positioning of an item on the shelf and the number of allocated facings. In this way, we obtain information on the importance of the shelf layout and the parameter estimates for the individual SKUs will be shrunk toward reasonable values, thereby dampening some of the undesirable variation that separate, independent estimators could have. The marketing literature contains many papers using hierarchical models (see, for example, Blattberg and George 1991, Montgomery 1997, and Boatwright et al. 1999). These papers document that the hierarchical model reduces the problem of coefficient instability across equations and that it improves predictive power.

Based on this model, we develop an optimization procedure for shelf management using SA. In contrast to existing literature, we explicitly account for a moderating effect of shelf layout on marketing mix elasticities. Furthermore, instead of restricting the analysis to shelf space, we also consider other shelf descriptors such as the horizontal and vertical position of an item on the shelf. Moreover, we develop our model for a large number of items. Instead of considering

the market at the brand level, we consider individual SKUs. Finally, we appropriately take into account uncertainty in sales and in the model parameters. We believe that this situation comes close to actual practice.

To summarize, our modeling and optimization approach is in various ways related to previous papers in the literature. We extend most previous shelf optimization approaches in at least one out of four important ways: (1) we account for dependencies between shelf characteristics and marketing-mix elasticities, (2) we use a rich description of the shelf layout instead of focusing on shelf space, (3) we optimize the layout while taking into account the uncertainty in sales and model parameters, and (4) we present a model that can easily be considered for a large number of items. Table 1 provides an overview of the present literature and its most important features.

3. Our Approach

The model proposed here aims to accurately measure the effect of shelf space and shelf placement on sales levels and on marketing instrument effectiveness. Furthermore, we avoid unrealistic simplifications in the shelf optimization.

The basis of our model is a standard sales equation that explains (log) sales from item- and time-specific intercepts and marketing effect parameters. We augment the sales equation with a second equation that relates the intercepts and effect parameters to shelf and SKU descriptors. We also control for the endogeneity that is likely to be present for the number of facings the retailer assigns to an item. Items that sell well tend to get more facings. To estimate the parameters of our two-level HB model, we use Markov Chain Monte Carlo (MCMC) methodology, in

Table 1 Papers on Shelf Layout Optimization

Paper	Shelf descriptors	Moderated variables	Estimated shelf effects? ^a	Optimization method	Maximum number of items in application
Corstjens and Doyle (1981)	Shelf space	Baseline sales	Yes	Geometric programming	5 (Product groups)
Corstjens and Doyle (1983)	Shelf space	Baseline sales	Yes	None	4 (Fictitious products)
Bultez and Naert (1988)	Shelf space	Baseline sales	Yes	Nonlinear programming heuristic	20
Drèze et al. (1994)	Shelf space & shelf location	Baseline sales	Yes	Integer programming	235
Borin et al. (1994)	Shelf space	Baseline sales	Yes	SA	18
Borin and Farris (1995)					
Yang and Chen (1999)	Shelf space & shelf number	Baseline sales	No	Nonlinear integer programming	6
Yang (2001)	Shelf space & shelf number	Baseline sales	No	Heuristics + knapsack	10
Lim et al. (2004)	Shelf space & shelf number	Baseline sales	No	Tabu search, Squeaky Wheel Optimization	100
This study	Shelf space & shelf location	Baseline sales & marketing elasticities	Yes	SA	407

^aIndicates whether the effects of shelf layout on sales are estimated, as opposed to assumed known as is often done in the optimization stream of the shelf optimization literature.

particular, Gibbs sampling. We use graphs to display the potentially complicated nonlinear effects captured by our model. To investigate how well the model performs in describing and forecasting the sales (and thus profits in the shelf optimization), we predict sales for a holdout sample of five weeks of data. A comparison of our forecasts to actual and forecasted sales using SKU-level regressions provides information on the absolute and relative performance of our model.

Next, we optimize the total profit for the final week by changing the shelf arrangement. Using the posterior draws of the model parameters from our Gibbs sampling algorithm, we obtain the probability distribution of sales and profit in the optimization period for any feasible shelf arrangement. We evaluate the shelf allocations using the expected value of the posterior gross profit distribution. Note that this measure gives the expected profit over all sources of uncertainty, that is, uncertainty in sales as well as in the (estimated) parameters. We set the minimum number of facings for each item to two. In this way we do not allow items to have zero or one facing(s), that is, we do not consider assortment decisions and move away from low values for facings, as we observe very few items with only one facing in our data. Assortment decisions are beyond the scope of our research. Item deletions and new product acceptance are driven by other factors such as inventory costs, retailer power, and retail chain positioning (see Sloot et al. 2006 and Luo et al. 2007).

There are many ways to search for the optimal shelf arrangement. Given the complexity of the problem, an algorithm that yields a guaranteed optimal solution is hard, if not impossible, to obtain. Therefore, we opt for SA, which is a heuristic approach, to search for the optimal shelf arrangement. The algorithm starts with a random shelf arrangement and then searches the neighborhood of the current solution for better ones. To avoid getting stuck in local maxima, an inferior solution may also be temporarily accepted, but the probability of this event decreases as the algorithm proceeds, see Johnson et al. (1989).

4. A New Model for Shelf Management

In this section we describe our modeling and optimization approach in detail. In §4.1, we discuss the model. Technical details on the estimation of the model parameters are presented in the appendix. In §4.2, we describe how we optimize the shelf layout.

4.1. Representation of the Sales Model

First, we introduce some notation. We denote the number of SKUs in the market by I , the number of observations for SKU i by T_i , and the number of item

attributes by L . Among the L attributes, there are $C \leq L$ shelf characteristics such as the number of facings, the shelf number, and the distance to the end of the shelf. To explain sales, we have K explanatory variables, such as price and promotion. Let $\ln S_{i,t}$ be the natural log of sales of SKU i at time $t = 1, \dots, T_i$. We model the sales by a standard log-linear model (see, for example, Wittink et al. 1988), that is

$$\ln S_{i,t} = X'_{i,t} \beta_{i,t} + \varepsilon_{i,t}, \quad i = 1, \dots, I, t = 1, \dots, T_i, \quad (1)$$

where $X_{i,t}$ denotes a $(K + 1) \times 1$ dimensional vector containing an intercept, the K explanatory variables for SKU i at time t , and where $\beta_{i,t} = (\beta_{0,i,t}, \dots, \beta_{K,i,t})'$ with $\beta_{k,i,t}$ the coefficient measuring the effect of the k -th explanatory variable for SKU i at time t . The vector of explanatory variables will, in general, contain (log-transformed) marketing instruments such as price, feature, and 0/1 dummy variables such as promotion. To allow for dynamic effects and if data availability permits, one may also include lagged prices and/or lagged sales (see Fok et al. 2006). We let the error term $\varepsilon_{i,t}$ be independently distributed $N(0, \sigma_i^2)$.

All β parameters in Equation (1) potentially differ across SKUs and across time. To describe how these parameters vary across these two dimensions, we add a second layer to the model. In this layer, we specify a model for the marketing effect parameters. As explanatory variables in this second-level model, we use item-specific attribute data and shelf layout descriptors, both of which may, in general, vary over time. Of these attributes, the shelf characteristics are most likely to change. This can happen for example because of a relocation of items during the observational period. We denote the observed attributes of item i by an $(L + 1) \times 1$ vector $Z_{i,t}$. This vector contains an intercept, the SKU characteristics, and the shelf layout characteristics at time t .

We introduce the following linear relation between the item-specific parameters and the attribute space, that is

$$\beta_{i,t} = \gamma Z_{i,t} + \eta_i, \quad \eta_i \sim N(0, \Sigma_\eta), \quad (2)$$

where γ is a $(K + 1) \times (L + 1)$ matrix of parameters. The coefficients $\gamma_{k,l}$ represent the effect of attribute l on the effect size of marketing instrument k . For $k = 0$, the coefficients represent the effects of the SKU characteristics on the intercept of the sales equation. In other words, these coefficients give the direct effect of the shelf layout on sales. Note that the current literature usually restricts the analysis to only these effects (Corstjens and Doyle 1981, Yang and Chen 1999, Yang 2001, Lim et al. 2004). Also, most papers only use facings as a shelf layout descriptor.

In §5, we show how we augment $Z_{i,t}$ with functions of shelf layout characteristics to allow for nonlinear effects, such as a decreasing elasticity of facings. In principle, one could also add lagged values of the shelf characteristics to capture possible dynamic effects of changing the layout. However, as changing the shelf layout is rather costly, we are now less interested in the transitory effects of a change in shelf layout. Furthermore, for most empirical cases, one will only observe very few changes. This makes the estimation of this type of dynamic effect very complicated and it is therefore not pursued here.

Of course, there may be relevant attributes that we do not observe, or intangible attributes such as brand equity that also influence baseline sales and the marketing instrument effectiveness. We represent the joint effect of such attributes by a normally distributed disturbance term in (2), that is, $\eta_i = (\eta_{i,0}, \eta_{i,1}, \dots, \eta_{i,K})' \sim N(0, \Sigma_\eta)$. Note that we assume that these intangible characteristics are fixed over time. This implies that we assume that a relocation of the products will not affect η_i . Because the degree of uncertainty may differ across instruments, we allow the variance of $\eta_{i,k}$ to depend on k . Furthermore, we may expect that some unobserved attributes simultaneously affect multiple marketing instruments. For example, if an item has a high feature effectiveness, it may also be very effective with display. Such relations will lead to positive correlations between $\eta_{i,k}$ and $\eta_{i,h}$, $k \neq h$. To capture such correlations, we allow Σ_η to be nondiagonal. Note that as our model is multiplicative, the absolute effect of shelf characteristics on sales depends on other observed characteristics as well as on η . Thus, we implicitly take into account the fact that different SKUs will have different facing elasticities.

An alternative view on Equation (2) is that an SKU can be represented by a specific point in an attribute space. The second layer of our model specifies a (linear) mapping from the attribute space to the model parameters in (1). Furthermore, by explicitly recognizing that items that are close in attribute space will also have similar parameters, we efficiently make use of the data to estimate marketing effectiveness parameters.

In sum, the combination of Equations (1) and (2) generates our attribute-based sales model. The joint estimation of the two equations gives more precise estimates of the attribute mapping than a two-step approach, in which Equation (1) would be estimated separately per SKU and where the resulting estimates of β_i would then be regressed on SKU and shelf characteristics. Our HB approach yields more accurate estimates as it combines all the available information and accounts for uncertainty in estimates of the marketing instrument effectiveness. Furthermore, in a

two-step approach it would be difficult to deal with changes in characteristics of shelf allocation. In the appendix, we discuss an MCMC algorithm that can be used to estimate the model parameters and which, as a by-product, gives draws from the distribution of all parameters conditional on the data.

4.1.1. Endogeneity. There is one potentially important problem with the model presented above. In practice, items that sell well tend to get many facings (see for example van Dijk et al. 2004). A manager is likely to use the baseline sales as a basis to construct the shelf layout. Obviously there are feedback loops and endogeneity in play here. Fortunately, for our data, halfway through the sample, the shelf layout is reorganized according to a specific experimental setting. Therefore the endogeneity problem may not be too severe. However, the layout in the first part of the sample is still set by the retailer.

If the endogeneity is ignored, we may overestimate the effect of facings on sales. In technical terms, endogeneity means that the number of facings (an element of Z_{it}) is correlated with the unobserved component driving the baseline sales of the item (the element of η_i corresponding to the intercept). To account for this endogeneity, we also model the retailer's decision rule for setting the number of facings. The most important determinants of the number of facings will be the unobserved components in η . Therefore we model the facings using

$$\text{facings}_{it} = \delta + \eta_i' \theta + \xi_{it}, \quad (3)$$

where $\xi_{it} \sim N(0, \sigma_\xi^2)$ denotes the unexplained part of the number of facings of item i at time t . The parameter θ specifies how the number of facings is influenced by the unobservable components in η_i . By including this relationship in the model, we correct for the possible endogeneity of the number of facings. Note that we can therefore safely assume that ξ_{it} is uncorrelated with all other components of the model. The relative importance of the endogeneity can be evaluated using the estimates of θ .

However, in Equation (3), we actually assume that the number of facings changes every time period. In practice, the shelf layout tends to be fixed over long periods of time. Assuming that the timing of a change in the layout is exogenous, we need only consider changes in the layout. We then obtain the following specification:

$$\text{facings}_{it} = \begin{cases} \text{facings}_{i,t-1} & \text{if no reallocation} \\ \delta + \eta_i' \theta + \xi_{it} & \text{if reallocation.} \end{cases} \quad (4)$$

If the layout changes only once in the data set, we actually have two settings for facings to model

for each item, that is, the number of facings before the reallocation and the number of facings after the reallocation.²

The full final model now reads as

$$\begin{aligned} \ln S_{it} &= X'_{it}\beta_{it} + \varepsilon_{it} \\ \text{facings}_{it} &= \begin{cases} \text{facings}_{it-1} & \text{if no reallocation} \\ \delta + \eta'_i\theta + \xi_{it} & \text{if reallocation.} \end{cases} \quad (5) \\ \beta_{it} &= \gamma Z_{it} + \eta_i. \end{aligned}$$

The parameters in the facings equation can easily be sampled along with the other model parameters, see the last two paragraphs of the appendix for more details.

4.2. Shelf Optimization

The output of the Gibbs sampling algorithm allows us to draw inference on the posterior distribution of any function of the parameters. The total profit of the category for a particular week and store is one example of such a function. Our model contains an extensive set of shelf arrangement descriptors in the number of facings, shelf height, and distance to the end of the shelf, described by Z . We can therefore obtain the posterior distribution of the profit based on the current shelf layout, as well as based on any feasible alternative layout, which is key to our approach. To calculate the expected profit value for a particular layout, we average over the parameter draws obtained in our MCMC algorithm. Note that the posterior profit distributions are conditional on the (in sample) data and they represent both the uncertainty in the sales levels themselves and the uncertainty in the parameters. In turn, we can use these distributions to optimize the expected total profit for out-of-sample periods, conditional on the data. In the end, we will obtain (only) one optimal layout for which we can inspect the posterior profit distribution.

As said, in this paper, we use the mean of the posterior profit distribution to measure the quality of the associated shelf arrangement. As an alternative, one could also consider the mode, or even the 5% percentile of the profit distribution. The latter would correspond to maximizing the profit under a worst-case scenario.

²Equation (4) should be seen as an approximation of the true model. In the end, the number of facings is a count variable. In Equation (4) this special property is not accounted for. However, we are not so much interested in the actual model explaining the number of facings. We include Equation (4) in the model to account for the (potential) correlation between the random effects η_i and the number of facings. An approximate model will suffice for this.

4.2.1. The Shelf Optimization Problem. Let $S_i(Z_i)$ be the sales for item i , given its shelf allocation and item characteristics Z_i . Let m_i denote the per unit contribution for item i and let $c_i(Z_i)$ be the replenishment cost for carrying item i for a given layout Z_i . This, in turn, also depends on the expected sales given layout Z_i , as more sales means more replenishment activity. Furthermore, define Π as the total expected profit for the category, that is

$$\Pi = \sum_{i=1}^I (m_i E[S_i(Z_i)] - E[c_i(Z_i)]). \quad (6)$$

Note that we have to take the expected value over both the sales and costs. The issue of interest is to maximize Π given several restrictions. The main restrictions concern the logical consistency of the shelf layout.

Formally specifying and solving a mathematical model for this problem is difficult. First, formal mathematical restrictions that correspond to the consistency requirements are difficult to formulate. Previous papers that did specify formal mathematical restrictions only considered the number of facings (and sometimes the shelf number) as a decision variable, while the exact location on the shelf was not taken into account. In this case, the restrictions are much easier to specify in a mathematical programming format.

Second, given the enormous number of possible combinations of facings, shelf numbers, and the other decision variables, it is impossible to find a closed-form solution for this optimization problem, in particular, if the number of SKUs is large. The geometrical programming framework, or branch-and-bound procedure, as used by Corstjens and Doyle (1981), would also have a hard time finding an optimal solution in the high-dimensional space.

In this paper we suggest a heuristic optimization technique to search for the profit-maximizing shelf layout in a practical retailer situation. In this case we opt for SA. One of the advantages is that with this method, we are sure that at any step of the optimization procedure we have a feasible solution. Even with more shelf descriptors, all restrictions are easy to check in practice, for example, (1) each SKU must be assigned to a shelf, and (2) the total width of shelf space used by items may not exceed the total shelf space available; and (3) the shelf space allocated to a particular SKU may not (partly) overlap with another SKU. If one would optimize the layout using, for example, linear programming, all these restrictions would have to be translated into formal mathematical equations using the decision variables. In the SA method, we approach these restrictions in a different way. In searching for the optimal shelf layout,

we consider only feasible layouts, thereby ensuring that the layout always satisfies the given constraints. In the next section, we present more details of this technique.

If needed, additional restrictions can easily be added for the particular retailer's situation. For example, it may be interesting to add restrictions on the capacity of the shelf space allocated to SKUs. In some cases, the capacity of the allocated shelf space must at least be equal to the minimum packout. That is, in case of restocking of the item, one full packout has to fit on the shelf. Incorporating such a restriction in our optimization strategy is very simple. We, again, just have to make sure that we do not consider layouts that violate these restrictions.

4.2.2. SA Applied to Shelf Optimization. SA was proposed by Kirkpatrick et al. (1983). One of the advantages of this algorithm is that in each step a feasible solution is guaranteed. In our setting this means that the layout in each iteration will comply with all logical consistency restrictions. In each iteration of the algorithm, new feasible layouts in the neighborhood of the current solution are considered. If a candidate solution performs better than the current one, the current solution is discarded in favor of the candidate. With SA, an inferior candidate solution may also be accepted, but this happens with a certain probability, which decreases with the difference in profit between the two solutions and it also decreases as the algorithm proceeds. In the terminology of the SA algorithm, this probability depends on the so-called temperature of the system, which decreases as the algorithm progresses. By allowing for the acceptance of inferior solutions, the algorithm lowers the probability of becoming trapped at local optima. At the end the final solution is the best candidate solution found during the progress of the algorithm. We refer to Johnson et al. (1989) for a more detailed description of SA, and to Borin et al. (1994) and Silva-Risso et al. (1999) for marketing applications.

For our shelf optimization problem, we let the SA algorithm start at the best of many randomly generated layouts. We generate a preset number of layouts at random and choose the one that has the highest predicted profit as the starting point. One can generate as many layouts as desired. This makes it (even) less likely for the algorithm to get trapped in local optima.

In the search for a neighborhood solution, we use two methods. The first generates a new layout by interchanging two randomly chosen SKUs, as far as their shelf height and position on that shelf are concerned. The number of facings for each SKU is then adapted upward or downward according to the space available in the new location. The second method randomly selects a shelf, and on this shelf, randomly

selects two SKUs. If feasible, the first SKU loses one facing while the other gains one. If this does not work, for instance, when the first SKU is already at the minimum number of facings allowed, the other way around is tried, that is, the first SKU gains one facing while the other loses one. If this is also not feasible, a new shelf and set of items are randomly drawn. As items may have different package widths, an extra check here is needed to make sure the items still fit on the shelf. If not, the gaining item loses its extra facing again. By searching the space in this way, we use the smallest step size available. Larger steps would involve interchanging several items at once, or using larger facing increases and decreases. Although computation time increases, we prefer small steps, as it prevents overlooking potentially promising solutions.

Note again that this optimization procedure is performed after obtaining the posterior distribution of all parameters through our MCMC algorithm. The obtained posterior draws are used to calculate the expected profit for each layout considered.

5. Application

To illustrate our method we present a detailed analysis of an interesting and extensive data set. In §5.1, we briefly describe the data. Section 5.2 presents the estimation results and the forecasting performance of our model. In §5.3, we illustrate how our model can be used to optimize the shelf layout in each of the stores in our data set.

5.1. Data Description

The data analyzed in this paper is a scanner data set with the sales levels of canned soup. The data concerns one of the categories studied by Drèze et al. (1994). The experiments in this study were carried out at Dominick's Finer Foods, a leading supermarket chain in Chicago. Sixty stores participated in the tests, where each store was randomly assigned to a control or test condition. There were two test conditions—*space to movement*, where the shelf sets are customized based on store-specific movement patterns, and *product reorganization*, where product placement is manipulated to facilitate cross-category merchandizing or ease of shopping. We choose to analyze the canned soup category because it has a large number of items and shows relatively frequent price changes. Furthermore, this category has large variation in shelf layout as one of the test conditions was to alphabetize the items on the shelf according to flavor.

In our analysis, we only look at stores that have data in the test condition. We have 36,044 observations for 407 canned soup SKUs, for five randomly selected test stores. Three stores carry 81 each and two carry 82 each of these items. There may be overlap between these items, but we treat them separately

Table 2 Available Variables in Attributes Equation

Variable	Description
Facings	
Facings	Number of facings on shelf in units
ln(facings)	Log of number of facings on shelf in units ^a
Vertical measures	
Shelf number	Shelf number (1 = bottom shelf, 5 = highest)
(Vertical distance to middle) ²	Square of distance of shelf variable to the middle shelf ^b
Horizontal measures	
Distance to shelf end (racetrack)	Number of inches to the racetrack
Horizontal distance to middle	Number of inches to the middle of the shelf
Item characteristics	
Item width	Width of the can in inches
Campbell	Item is from the Campbell brand (1 = yes, 0 = no)

Note. More variables were available in the data set, but these were removed because of too much correlation with the variables in this table.

^aIncluded to model diminishing effects.

^bSquared version turns out to capture the diminishing effects better compared to the absolute distance.

because items in different stores will differ in their position on the shelf. Even if an item would have the same location, it is unlikely that it will have the same demand and elasticity parameters. For each SKU, we have around 100 weeks of observations. As explanatory variables in the sales equation, we use an intercept, price, and a promotion variable, which is a combination of the variables bonus buy and display available in the database.

The unique feature of this data set is that we have information on a number of shelf characteristics and item attributes. We list these attributes in Table 2. The list of variables in the data set is longer but because several of these variables appear to correlate strongly with each other, some were removed, so as to obtain a list of managerially relevant variables. Following most papers cited earlier, the number of facings is an important determinant of demand. We also include the log of facings as a explanatory variable to capture the potentially diminishing effects of the number of facings on marketing instrument effectiveness.³

Another important descriptor that we use is the shelf number. We expect that products that are higher

³ The effect of facings might actually follow an S-shape. However, we expect that if increasing returns to facings occur, this is only the case for a very small number of facings (say, 1 or 2). Considering that we only have a few items with 1 facing, it would be very difficult to identify such an S-shape. Therefore we assume that the facings show diminishing returns. Furthermore, in our optimization, we restrict the minimum number of facings per SKU to 2. Finally, note that Drèze et al. (1994) did not find evidence of an S-shaped effect for this particular category.

Table 3 Descriptives of Selected Variables in the Data Set

Variable	Per item averages		Per item st. dev. Average ^b
	Average across items	St. dev. ^a across items	
Sales	323.75	650.59	194.64
Price (\$)	2.65	0.75	0.22
Promotion (%)	5.00	7.00	16.00
Shelf	3.43	1.17	0.38
Facings	4.15	2.93	0.46
Campbell (%)	87.00	33.00	0.00

Note. Both variation across items and across time is shown.

^aThis value gives an indication of variation across items.

^bThis value gives an indication of variation across time per item.

on the shelf have higher visibility. At the same time, some decreasing returns of shelf height may also appear. To capture this, we use the distance to the middle shelf (in our case the third shelf) as a moderating variable. We also use the distance of an item to the end of the shelf as a shelf descriptor. Products that are closer to the beginning of the shelf may benefit from people reaching the item quicker coming from the back aisle, or the “racetrack” (Larson et al. 2005, p. 395). On the other hand, we may find the opposite effect in that items that are in the middle get more attention from consumers, who may often end up in the middle of the shelf for the category. To capture these potential nonlinear position effects, we add the distance to the middle of the shelf to our set of shelf descriptors. Next to characteristics of the shelf layout, characteristics of the product itself, such as item volume and brand name, are incorporated because we expect they will influence demand and elasticities.

If the available data would allow it, one can extend this list with additional item characteristics, such as flavor, type (condensed or not), and package type (for example, Easy Open lid or not). These variables could contribute to the explanatory power of our model and yield additional insights.

We display some descriptives statistics in Table 3. The numbers show that sales vary quite a lot across items and across time. For price, we see that there is quite a lot of variation across items but a lot less within items.⁴ For facings and shelf, we see that, in general, there is relatively little variation within items over time.

5.2. Estimation and Forecast Results

In this subsection, we discuss the estimation results for our model, and report the out of sample forecast performance.

⁴ Because of this, we have to abstain from modeling dynamic effects of prices on sales.

5.2.1. Estimation. As explanatory variables in the sales equation we use an intercept, log price, and promotion. Each parameter associated with these variables is item and time specific. In the model, changes in the parameters across time for a specific item are completely attributed to changes in its characteristics. Obviously, differences across items can only partly be explained by differences in characteristics. The random component in Equation (2) allows for unexplained differences in the parameters across items. Shrinkage estimation in the hierarchical structure allows us to estimate these parameters with sufficient accuracy. For estimation of the parameters, we generate 20,000 iterations of the Gibbs sampler for burn-in and 20,000 iterations for analysis. We retain every tenth draw to reduce the effects of autocorrelation between consecutive draws. The (unreported) iteration plots are inspected to see whether the sampler has converged.

The marketing effectiveness parameter $\beta_{i,t}$ varies across items and time. Even though the values do not change every period, there are obviously too many values to display in a table. A histogram per marketing instrument, as given in Figure 1, insightfully

summarizes the dimensions item and time. Obviously, there is also some uncertainty around each of the points that we cannot display here. The number of observations that constitute the histogram is equal to $\sum_{i=1}^I T_i = 36,044$. We obtain the expected signs for each of the three explanatory variables. The intercept is positive for all observations. The price effect is negative for 85% of the periods and items. When it is positive, it is not significant all the time (2,264 out of 5,514 are significant). We think these positive numbers stem from the fact that the data do not contain substantial price variation for all of the items.

Finally, the promotion variable has the expected positive effect for most items and periods.

From Figure 1, it may be hard to see what the actual expected β parameter values are across all items in the data set. We therefore list the posterior means of the averaged β across all items and time periods ($(1/I) \sum_i (1/T_i) \sum_t \beta_{it}$) in Table 4. From this table, it is clear that across all items and time periods the parameters have the expected signs. To illustrate the changes that the β parameters experience when the layout changes, we include the averages of the changes in the β parameters in the righthand column.

Figure 1 Histogram Per Marketing Instrument Across All Items and Time Periods

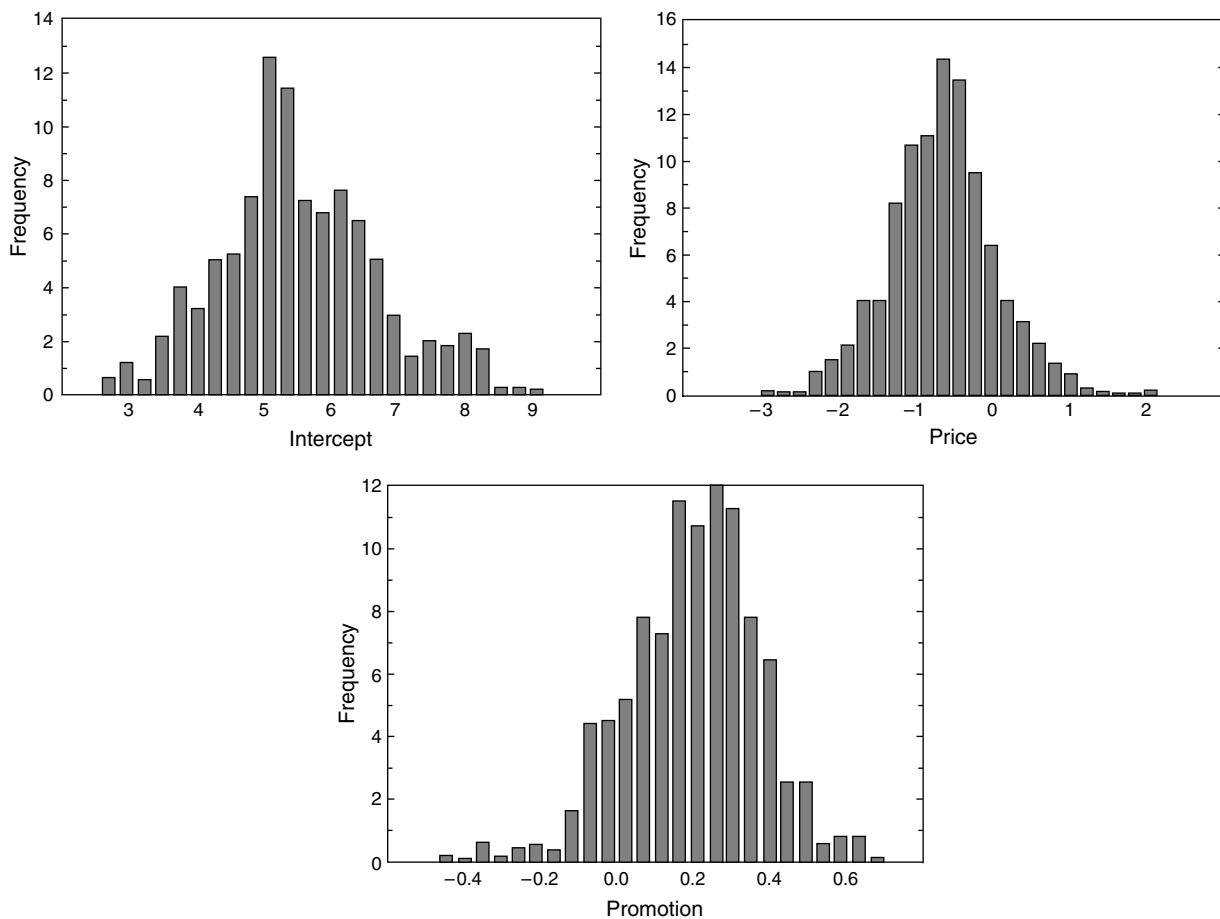


Table 4 Posterior Means (and Standard Deviations) for $(1/I) \sum_i (1/T_i) \sum_t \beta_{it}$

Variable	Mean	St. dev.	Average change after layout change
Intercept	5.502	0.079	-0.1555
Price	-0.658	0.065	0.1077
Promotion	0.195	0.027	0.0376

Before discussing the estimates of the parameters linking the shelf characteristics to the marketing effect parameters (γ), we examine the estimates for the facings equation. Recall that through this equation, we correct for the possible endogeneity of the facings. These parameter estimates are displayed in Table 5. Note that we find evidence that the number of facings is set endogenously. There is a significant relation between the random effects (η_i) and the number of facings, particularly with respect to the intercept and price. The items that have a high baseline sales, after accounting for marketing instruments and shelf characteristics, tend to have many facings. Items that have a relatively strong price elasticity, after accounting for all observed variables, tend to get fewer facings. Promotion sensitivity appears to have no effect on facings.

Table 6 shows the posterior means for γ , that is, the parameters linking the attributes to the effectiveness of own marketing instruments. In a nonlinear fashion, we investigate these estimates with the numbers in the table and with graphs for the attributes that have a nonlinear effect, i.e., facings, shelf, and distance to shelf end. From the numerical estimates, it can be seen that the logarithm of the number of facings has a positive influence on the intercept in the sales equation (0.455). This indicates that items with many facings have a higher expected baseline sales, that is, for the case where there is no promotion, feature activity, or otherwise. This is the effect that has been studied in the previous literature on shelf management. The number of facings appears to make the price effect stronger. However, there seem to be decreasing returns of facings with respect to the price elasticity.

Table 5 Posterior Results Endogeneity Parameters

Variable	Mean ^a	St. dev.
Intercept (δ)	1.287***	0.029
Effect of random effect, η , on facings (θ)		
Baseline sales (η_{i0})	0.421***	0.032
Price elasticity (η_{i1})	0.358***	0.050
Promotion effect (η_{i2})	0.099	0.121
σ_ξ^2	0.412***	0.030

Note. See Equation (4).

^a.,***: Zero not contained 99% highest posterior density region.

Table 6 Posterior Means (and Standard Deviations)^a for γ

Variable	Intercept	Price	Promotion
Intercept	7.537*** (0.982)	-0.262 (1.144)	1.258*** (0.472)
Number of facings	-0.025 (0.019)	0.093*** (0.020)	-0.020 (0.017)
Log number of facings	0.455*** (0.111)	-0.524*** (0.110)	-0.030 (0.110)
Shelf number (1, ..., 5)	0.136*** (0.025)	-0.175*** (0.030)	-0.068*** (0.025)
Squared vertical distance to middle shelf	-0.057*** (0.014)	0.029** (0.020)	0.005 (0.014)
Horizontal distance to shelf end	-0.003*** (0.001)	0.003*** (0.000)	-0.002*** (0.001)
Horizontal distance to shelf middle	0.007*** (0.001)	-0.006*** (0.000)	-0.002* (0.001)
Item width in inches	-1.464*** (0.359)	0.303 (0.420)	-0.214 (0.149)
Campbell	1.106*** (0.236)	-0.404* (0.210)	0.122 (0.087)

Note. The numbers in the cells reflect the effect of a layout characteristic (left) on a marketing mix instrument (top)

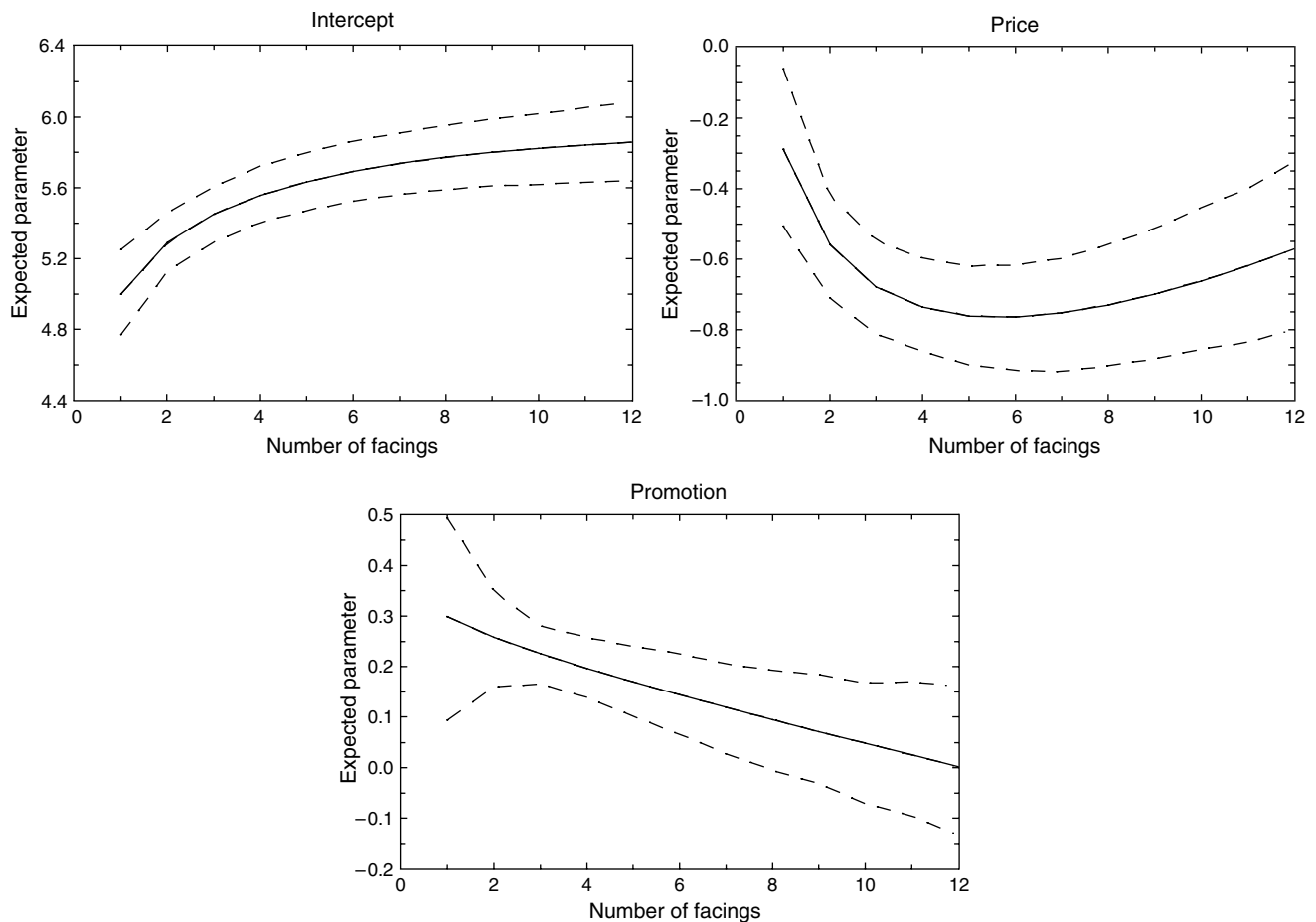
^a.,***,***Zero not contained in 90%, 95%, or 99% highest posterior density region, respectively.

Had we ignored the endogeneity, these parameters in particular would have been affected. To validate this, we ran a separate estimation where we do not account for the endogeneity, and find slightly stronger effects of facings on both the intercept and price elasticity. Probably the experimental nature of this data set helps to partly remove the endogeneity.

At this point, we could proceed with a model selection strategy to remove irrelevant relations between shelf or item descriptors and the marketing effect parameters. However, note that all variables are statistically relevant for at least one of the marketing effect parameters. We therefore cannot remove a variable from the model entirely. Furthermore, in the optimization, we take into account the uncertainty about all model parameters. More specifically, if we use the current model, we fully acknowledge that the nonsignificance of some elements of γ may stem from a lack of sufficient data. If we remove some relationships from the model, we actually presuppose the irrelevance of these elements.

To further investigate the effects of the two facing variables available in the model, we calculate the posterior mean of β for different values for the number of facings. The first graph of Figure 2 shows the effect of facings on the baseline sales (β_{i0}), that is, the direct effect of facings on sales. As discussed above, a higher number of facings causes a higher intercept, which, in turn, results in more sales. However, this effect levels off as the number of facings increases. As expected, the effect of the number of facings on sales exhibits diminishing returns. This effect is very similar to that found in Drèze et al. (1994). Next, in the second graph (top right), the effect of facings on the price elasticity is displayed. For a low number of

Figure 2 Impact of Different Number of Facings on Marketing Effectiveness Parameters



Note. Dashed lines show 95% highest posterior density region.

facings, the price elasticity appears to be higher than -0.5 . This could stem from the fact that these items are not very visible on the shelf, therefore the price is not easily observed by consumers. Price discounts, therefore, may only have a limited effect for these items. However, the price sensitivity increases as the number of facings goes up. Apparently, having more facings increases awareness of the price level, and thus creates more price sensitivity among customers. This effect appears to level off as the number of facings increases further. It even looks like price sensitivity reduces again as the number of facings increases beyond, say, 7 facings. However, taking the parameter uncertainty into account, we cannot base strong conclusions on this. This finding may be an artifact of our data, which show limited price variation over time for many items and contain relatively few items with this many facings.

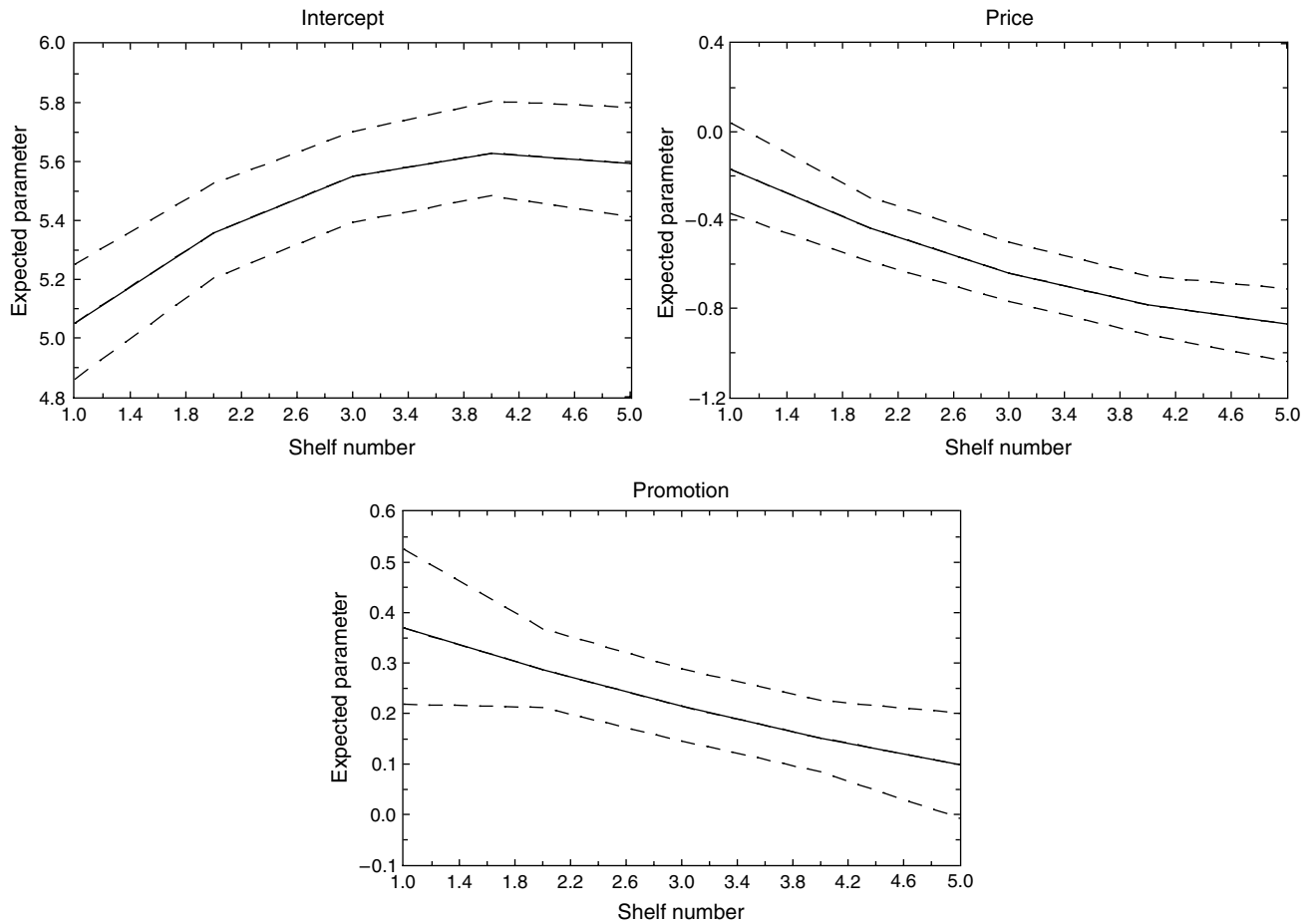
Finally, the third panel of Figure 2 shows the impact of facings on promotion effectiveness. Interestingly, items that have more facings have a slightly weaker promotion effect. For items that are already

very visible on the shelf, a promotion does not generate much additional attention.

Shelf height also seems to correlate positively with sales. This can be seen from the value of 0.136 for the effect of shelf on the intercept (see Table 6). Just as with facings, we see that a higher value (i.e., a higher shelf location) makes consumers more price sensitive. This is no surprise, as consumers see the prices for higher located products more easily than for those at the bottom shelf. This can also be seen from Figure 3. However, the effect on baseline sales levels off as we move toward the top shelf. Again this is consistent with the findings of Drèze et al. (1994). The other graphs show that the shelf number also affects price and promotion effects, which turn out to be close to linear.

The distance to the shelf end has a negative effect on sales, that is, the further away an item is from the racetrack, the lower the expected sales. The horizontal distance to the middle of the shelf has a small positive effect, so being further from the middle may increase sales. The combined expected effect can be seen in Figure 4. Obviously, being close to the racetrack is optimal. Note, however, that price sensitivity

Figure 3 Impact of Shelf Number (Varying from 1 to 5) on Marketing Effectiveness Parameters



Note. Dashed lines show 95% highest posterior density region.

is highest for these items as is promotional sensitivity. Being in the middle may hurt sales, although it makes consumers less price sensitive. Promotion effectiveness decreases as items move beyond the middle of the shelf. Surprisingly, Drèze et al. (1994) found no significant effect of the horizontal distance. However, they did not include the effect of the distance on price elasticity and promotion effect. As the graphs for the intercept and price in Figure 4 have the opposite shape, it is not unlikely that one finds no effect of the horizontal distance if the effect through price is ignored. Returning to Table 6, note that the brand Campbell's has higher expected sales and consumers appear to be more price sensitive for this brand.

Finally, Table 7 shows the posterior means for Σ_{η} . From these estimates, we conclude that there is quite a large proportion of the differences in baseline sales and price elasticities across the items that we cannot explain using item and shelf characteristics. Furthermore, there is a rather large correlation between the unobserved components of price and the intercept. SKUs with (unexplained) high baseline sales also tend to have a large price elasticity.

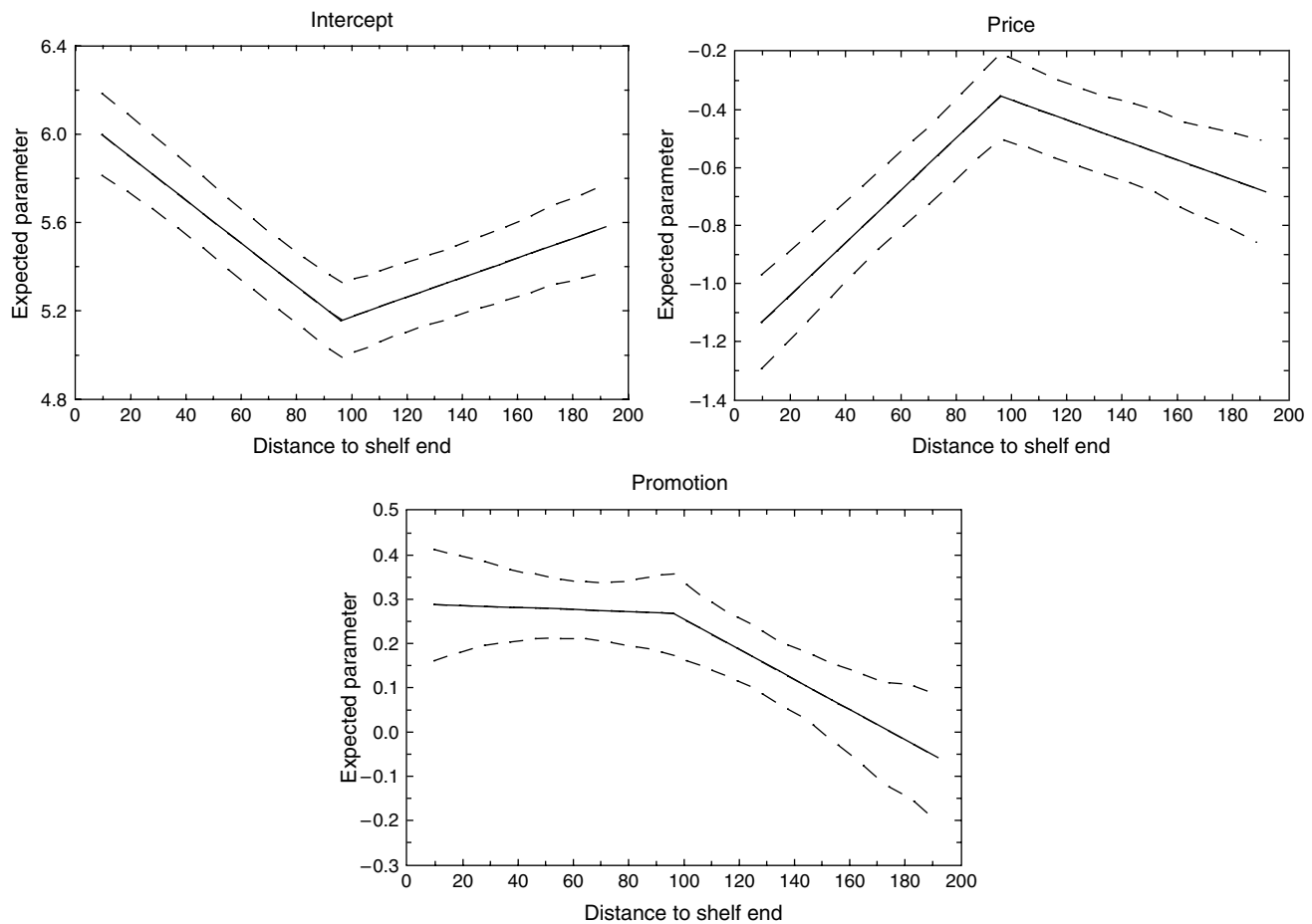
The parameter Σ_{η} is the only parameter for which we impose a (slightly) informative prior. Therefore it is interesting to compare the posterior mean of Σ_{η} with the mean under the prior. Under our prior (see the appendix), the expected value of Σ_{η} equals the identity matrix. The posterior mean for the variance corresponding to the price elasticity is relatively close to the prior mean. The same holds for the covariance of the random effect of promotion with the other two random effects. It may seem as if we have not learned much from the data. However, note that the standard deviations are much smaller than expected under the prior. So we have indeed learned a lot. We also observe that the other elements of the posterior mean of Σ_{η} are further away from the prior mean.

Table 7 Posterior Means (and Standard Deviations)^a for Σ_{η}

Mean	Intercept	Price	Promotion
Intercept	1.743*** (0.218)	-0.939*** (0.161)	-0.008 (0.048)
Price		0.859*** (0.128)	0.038 (0.029)
Promotion			0.068*** (0.011)

^a .*, **, ***Zero not contained in 90%, 95%, or 99% highest posterior density region, respectively.

Figure 4 Impact of Distance to Shelf End (or End of the Category) on Marketing Effectiveness Parameters



Note. Dashed lines show 95% highest posterior density region.

Table 8 Forecasting Results for Fifth Store in Data Set

Statistic	Full model (%)	Regression (%)
In sample		
Correlation	88.6	81.3
MAPE	10.5	11.6
Out of sample		
Correlation	89.4	60.9
MAPE	8.9	15.6

5.2.2. Forecasting. In each run of the Gibbs sampler, we simulate sales forecasts for the periods after the layout change for store 5. These data are not used for parameter estimation.⁵ We use the posterior mean of these forecasts as the out of sample prediction. We also predict the sales in sample. The forecasting diagnostics are presented in Table 8. The predicted sales correlate as much as 89% with true sales. The mean absolute percentage error (MAPE) statistics show average percentage errors of around 10% in

⁵ Although there are not many item deletions, store 5 had the most. Therefore it is a good candidate to be considered for out-of-sample data, so that parameter estimation would be affected the least.

sample and slightly lower out of sample, with 9%. We compare our model against a per-item regression and achieve better results for all diagnostics, in particular out of sample.

The real power of our new model however amounts to our ability to optimize the shelf arrangement, as we will see in the next subsection.

5.3. Shelf Optimization

5.3.1. Results for Five Stores. We perform the optimization of the shelf layout for each of the five stores separately. The SKU prices and promotion indicators are important determinants for the optimal layout. As the goal is to find a layout that performs well in general, we set these prices and promotion indicators equal to the average over the last 10 weeks of the data per store. In this way, we average over a number of different scenarios for the SKU prices and promotions. The replenishment costs $c_i(Z_i)$ in Equation (6) are currently assumed to be equal to 0.⁶

⁶ This setting can easily be changed. Experiments with different settings in the c_i -function showed close to identical results.

To begin the SA search process, we generate 10,000 random shelf arrangements and have the algorithm start at the arrangement that has the highest expected profit.

In addition to our profit-optimization routine, we compute profits for commonly applied rules of thumb. Because these rules do not completely prescribe the shelf layout, we generate 10,000 random layouts based on each rule at hand. The profit for the best of each of these 10,000 layouts is displayed in Table 9. Note that in reality, retailers could not do this, but would need a model like ours to predict which of many possible layouts would work best. More likely, they will develop additional rules of thumb to guide the placement of SKUs. Of course, these rules will not be optimal.

The first basic rule of thumb we consider is *share-of-shelf equals share-of-log-sales*. We have chosen the version of this rule with log-sales because in practice large items hardly ever get their share of sales in shelf space, and smaller items usually get more than their sales share justifies. The second rule of thumb is *share-of-shelf equals share of margin*. Retailers often devote more shelf space to products with high margins, rather than sell much. This rule does not appear to work very well, as will be seen later. Varying the position of high-margin items also has its consequences. It is to be expected that putting high-margin

items close to the beginning of the shelf results in higher profits. As shown by Larson et al. (2005), this is because shoppers do not always travel the entire aisle. In fact, once they enter an aisle, shoppers rarely make it to the other end. Instead, they travel by short excursions into and out of the aisle, which may lead them to purchase more from the beginning of the aisle than from the middle.

In Table 9, we present the optimization results. For the five stores in our data set, the SA algorithm finds profit increases relative to the current situation, that is, profit increases vary from 25% to 32%. Furthermore, the SA algorithm performs better than the rules of thumb described earlier, with increases ranging from 26% to 29%. Note that while the rules of thumb did not lead to a profit increase for stores 1, 3, and 4, our optimization method succeeds in finding better shelf layouts.

Contrary to what one might expect, items that gain profit do not necessarily have an increased number of facings. It may also be that the item is put on a different shelf, closer to the racetrack, or a combination of these things. In Figure 5 we display the number of facings an item had before and after optimization, for the first store in our data set. Indeed, some items that have fewer facings end up having more profit because of the effects of other important variables, such as distance to the racetrack and height. Apparently, it helps to be close to the racetrack, or on a higher shelf. As can be seen in the figure most of the items that had just one facing in the original data set are now assigned two facings. Note again that in the optimization we have chosen to restrict the range of values for facings to a minimum of two. In this way we ensure that the new layout is based on the part of the facings curve for which we have enough observations. Finally, in Figure 6 we show the layout for one of the stores before and after optimization, including the location of items before and after application of our SA algorithm. We give the following two interesting examples of insights that can be obtained from this figure.

(1) Item 50 was on shelf 3 with 4 facings and is now on shelf 4 with 11 facings. Not surprisingly, it experiences a large (expected) profit increase of 41%.

(2) Although item 77 had its facings reduced from 4 to 2, it still shows a profit increase because it was moved from shelf 5 to 4, which is a slight improvement (see Figure 3). Apparently, this offsets the move to the center of the category, which can be expected to be unfavorable (see Figure 4).

Before the optimization, the figure shows that most of the bulk items are at the bottom, whereas the top

Table 9 Profit Results for Current Layout, Various Rules of Thumb, and Optimization Algorithm

Layout	Resulting maximum profit				
	Store 1	Store 2	Store 3	Store 4	Store 5
Current layout ^a (\$)	3,506	2,308	1,812	3,147	2,632
Rule of thumb ^b					
Share of shelf = share of sales (\$)	3,359	2,362	1,783	2,957	2,679
Share of shelf = share of margin (\$)	3,449	2,381	1,798	3,059	2,634
Put high margin items close to racetrack ^c (\$)	3,371	2,386	1,785	3,037	2,606
Put high margin items far from racetrack ^c (\$)	3,293	2,301	1,750	2,858	2,523
Optimization ^d					
Optimized layout (\$)	4,429	3,023	2,263	3,872	3,463
Improvement over current layout (%)	26	31	25	23	32
Improvement over best rule of thumb (%)	28	27	26	27	29

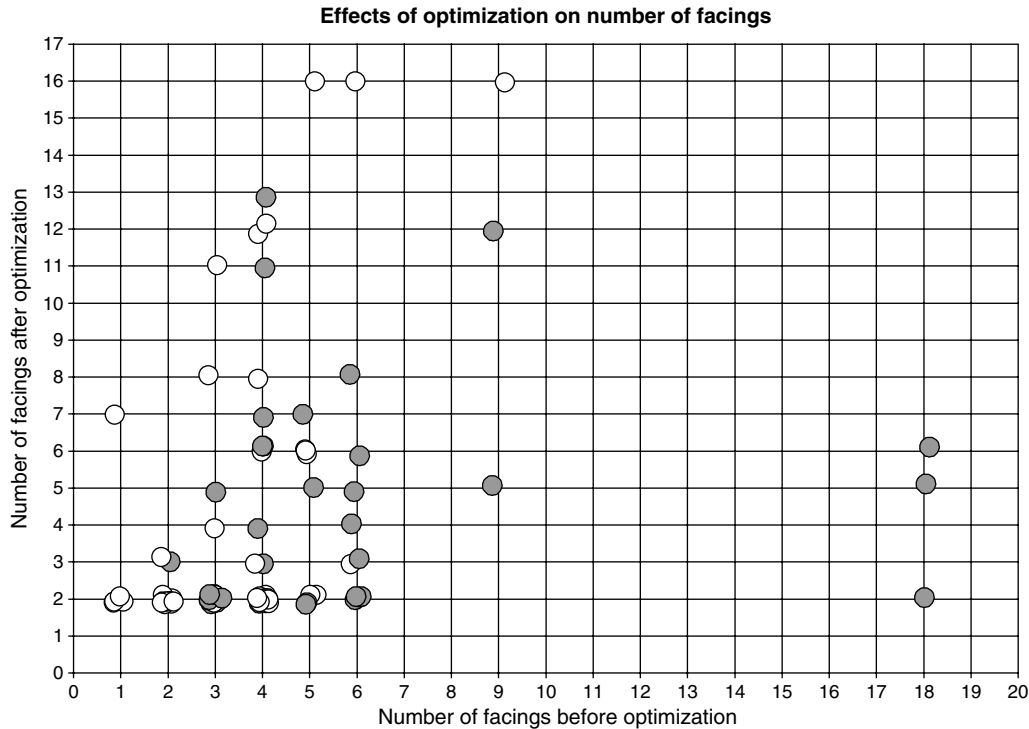
^aProfit based on predicted sales (not actual).

^bAchieved by generating 10,000 random layouts, where an item gets devoted the share of shelf space based on sales or margins. The profit for the best of each of these 10,000 layouts is displayed.

^cSame as "Share of shelf = share of margin," in addition, high-margin items are put close to beginning or end of shelf (where the beginning is the back-aisle or "racetrack").

^dThe profit for the optimized layout results after running our SA algorithm.

Figure 5 Number of Facings Before and After Optimization for Store 1

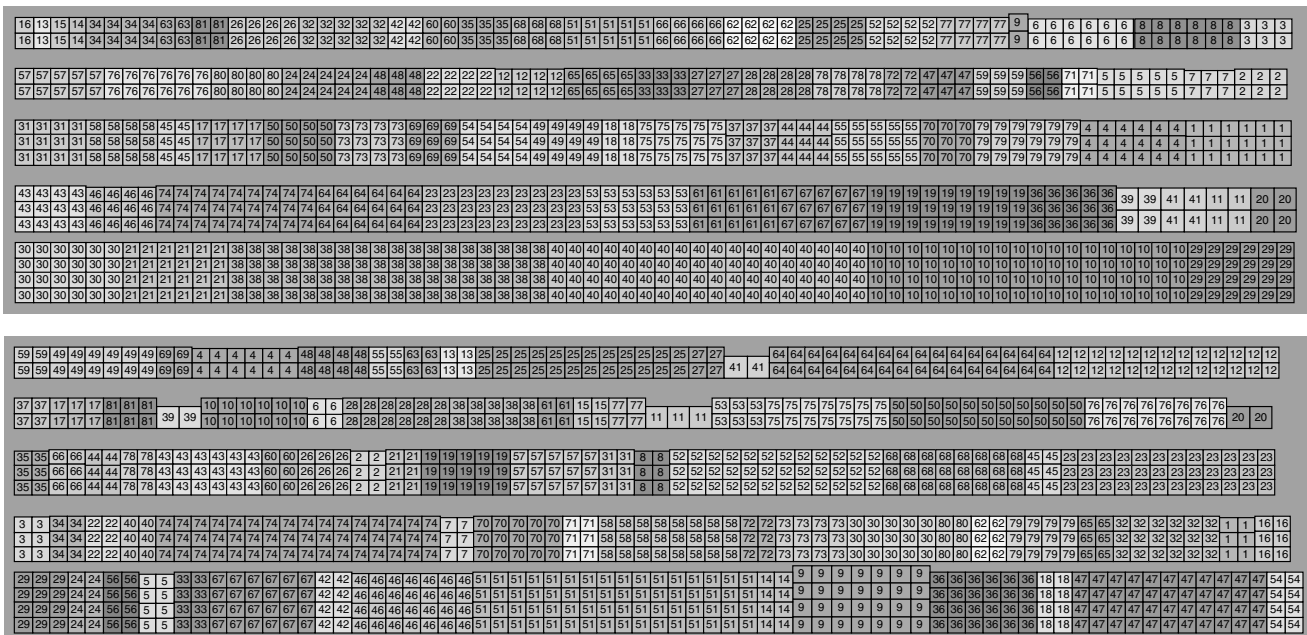


Note. Black bubbles reflect items that have reduced profit after optimization, white bubbles are used for items that have increased profit after optimization.

shelf seems to contain more specialty items. After the optimization, this is no longer the case across the board. Items with one facing have been assigned two facings, as this is the minimum in the optimization. The algorithm also appears to prescribe that certain items should move to the bottom, and get a low number of facings (specifically two, the minimum). Appar-

ently, it is beneficial to get low profits from these items, and leave more space on the other shelves for items that yield higher profits. However, if the retailer still prefers to have high facing items on the lowest shelf, perhaps for easier refills, we could add this as a restriction to the optimization by not considering layouts that suggest having few facings on low shelves.

Figure 6 Drawing of Optimal Shelf (Bottom) vs. Original Shelf (Top)



Copyright: INFORMS holds copyright to this Articles in Advance version, which is made available to institutional subscribers. The file may not be posted on any other website, including the author's site. Please send any questions regarding this policy to permissions@informs.org.

Table 10 Profit Increase for Varying Facings Restrictions

Maximum facings	Minimum facings (%)		Total ^a (%)
	2	3	
9	11.5	7.4	9.4
10	13.7	9.3	11.5
11	15.4	9.8	12.6
12	17.3	10.8	14.0
13	19.0	11.3	15.1
14	21.0	12.7	16.9
15	23.2	13.9	18.5
16	24.9	14.1	19.5
Total ^a	18.2	11.2	14.7

^aThe total column and row indicate the average across rows and columns respectively.

5.3.2. Restrictions. It may be expected that we would find smaller reported profit increases if more retailer-specific restrictions are built into the optimization problem. Our optimization algorithm can easily cope with these restrictions, by not considering neighboring layouts that violate restrictions. For example, we experimented with setting the minimum number of facings to two or three.⁷ A higher minimum number of facings of course leads to lower profits. If needed, the minimum number of facings can even be made item-specific. We also varied the maximum number of facings that could be assigned to a SKU. We report results in Table 10. The results show that for an increasing number of maximum facings, a higher profit increase can be expected. If we allow the optimization algorithm less freedom, by increasing the minimum number of facings from 2 to 3, we find much lower profit increases (11% versus 18%). The sensitivity with respect to the maximum number of facings is much smaller than with respect to the minimum. With an increasing maximum number of facings, profit increases seem to slow for the minimum of 3, but continue to increase for a minimum number of 2.

Although we also experimented with putting the private label at eye level (shelf 4 or 5) versus lower shelves, this did not have a large effect on profit levels, which probably stems from the fact that our data set does not contain that many private label items. We expect that if one has more private label items, this experiment may show larger effects on the expected profit increase. Of course experiments with other restrictions are also possible. In practice, the actual restrictions to be imposed will depend on the specifics of the particular situation.

⁷ Using four facings (an unrealistically high value) is not feasible because it gives very few degrees of freedom with, for instance, the rules of thumb.

6. Conclusion and Further Research

In this paper, we have presented a new approach to measure the impact of shelf layout on sales and to optimize shelf arrangements. By introducing shelf characteristics in an HB fashion into a sales model, we were able to model the direct effect of the shelf layout on sales as well as the moderating effect of the layout on the marketing instrument effects. We also took into account the endogeneity of facings. After estimating the model parameters on experimental data, we found that the shelf layout has significant effects on baseline sales and marketing effectiveness. This not only holds for the number of facings allocated to an item, but also for other shelf descriptors such as shelf height and distance to the end of the aisle. We also noted that, items that sell more appear to get more facings, as was picked up and controlled for by the endogeneity equation in our model.

Our HB setup facilitated interesting (graphical) insights into the effects of shelf layout on often-used marketing instruments such as price and promotion. Managerial implications derived from these graphs are the following: As expected, an SKU with more facings has higher sales. However, the additional benefits of one extra facing does decrease. Also, price effects appear to be weaker for items with few facings. Finally, the results show that promotion effects are weak when products have more facings or are located further away from the racetrack. These are important implications that could not have been derived from previously proposed sales and shelf management models.

Furthermore, our approach allowed us to optimize the shelf arrangement by cleverly searching through the huge dimensions of search space that any reasonably large sized category would provide. The SA algorithm found increases in profits for all stores in our data set. Optimized profits were also higher when compared to several rules of thumb. Sensible rules of thumb increase profits when compared to the current layout only for some stores. Our optimization technique allows for the identification of high potential SKUs that could give more profit to the retailer when put on the proper shelf location. This meant that certain low potential items were put on less favorable shelves.

Although the model presented in this paper could be extended in various directions, based on the amount of data per store and the little variation observed in prices the present data might not support further extensions. However, we suggest several directions for further research. It would be interesting to analyze different categories. On the technical side, one could consider specifications that accommodate spatial covariance, seasonality, and varying variance over time. If there is sufficient variation in the

observed data, one could combine shelf and price optimization. It would also be interesting to consider even more flexible functions for the nonlinear effects of the number of facings, the shelf number, and the distance of an item to the shelf end. If the data would allow it, one could estimate the effect for each available shelf. For the other two variables, one might consider a semiparametric approach using splines (see, for example, Sloot et al. 2006). The actual shape of the curve may prove to be important for the resulting layout, especially when, in the optimization, the minimum number of facings is set to one.

We see our model as a useful tool in analyzing the effects of shelf layout on marketing instrument effectiveness, optimizing shelf layout, and determining the value of SKUs to the retailer.

Acknowledgments

The authors, thank Alan Montgomery for providing the data, and Paul Farris, Kusum Ailawadi, and Richard Paap for useful discussions. The authors also thank the Area Editor and two anonymous reviewers for very helpful and detailed comments; and seminar participants at Carnegie Mellon University in Pittsburgh and Erasmus University Rotterdam, The Netherlands.

Appendix. Parameter Estimation

This appendix describes the algorithm for sampling from a Markov Chain with the posterior distribution of the model parameters as its stationary distribution (see Tierney 1994, and Casella and George 1992). In particular, we use the Gibbs sampling technique of Geman and Geman (1984) with data augmentation (see Tanner and Wong 1987). The latent variables η_i , $i = 1, \dots, I$ are sampled alongside with the model parameters. In our model, we define η_i as the latent variable of interest instead of $\beta_{i,t}$ as the changes in $\beta_{i,t}$ over time are deterministic. In the algorithm below, we consider the full model, that is, including the component dealing with the endogeneity of the number of facings. The sampling steps for the model without this component can easily be obtained by removing some elements from the discussion below.

Before we can present the likelihood function corresponding to the model in Equations (1), (2), and (4) we have to introduce some additional notation. Denote the number of observed configurations of the shelf layout by M , that is, if there were one reallocation of the shelf layout in the observed data, then $M = 2$. Denote by T_m the time index of the first observed period of each configuration m . Note that T_1 will always be 1. Using this notation, the likelihood equals

$$L(\text{data} | \vartheta) = \prod_{i=1}^I \int \left(\prod_{t=1}^T \phi(\varepsilon_{i,t}(\gamma, \eta_i); 0, \sigma_i^2) \right) \cdot \left(\prod_{m=1}^M \phi(\xi_{i,T_m}(\delta, \theta, \eta_i); 0, \sigma_\xi^2) \right) \phi(\eta_i; 0, \Sigma_\eta) d\eta_i, \quad (7)$$

where $\vartheta = (\text{vec}(\gamma)', \sigma_1^2, \dots, \sigma_I^2, \text{vec}(\Sigma_\eta)', \delta, \theta', \sigma_\xi^2)$ is the vector of all model parameters and

$$\begin{aligned} \varepsilon_{i,t}(\gamma, \eta_i) &= \ln S_{i,t} - X'_{i,t}(\gamma Z_{i,t} + \eta_i) \\ \xi_{i,T_m}(\delta, \theta, \eta_i) &= \text{facings}_{i,T_m} - \delta - \eta'_i \theta. \end{aligned} \quad (8)$$

The first line in Equation (8) corresponds to the error in the sales equation given the model parameters and the latent variable η_i , the second line corresponds to the error in the facings equation.

We impose flat priors on all parameters but the covariance of η_i . For this covariance we use an inverted Wishart prior. The full prior distribution equals

$$p(\gamma, \sigma_1^2, \dots, \sigma_I^2, \Sigma_\eta, \delta, \theta, \sigma_\xi^2) \propto \prod_{i=1}^I \sigma_i^{-2} \times f(\Sigma_\eta; \lambda, S), \quad (9)$$

where $f(\Sigma; \lambda, S)$ is the density function of an inverted Wishart distribution with λ degrees of freedom and scale parameter S evaluated at Σ . The value of λ is $K + 4$ and S equals $2\mathbf{I}_{K+1}$, where \mathbf{I}_m denotes an m -dimensional identity matrix. Although the influence of this prior on the posterior distribution is only marginal, the performance of the MCMC algorithm is significantly improved by imposing the inverted Wishart prior (see Hobert and Casella 1996).

Sampling of γ

After combining Equations (1) and (2) and stacking the sales equations over t , we obtain

$$\ln S_i = X_i^{*'} \begin{pmatrix} \gamma Z_{i1} \\ \gamma Z_{i2} \\ \vdots \\ \gamma Z_{i,T_i} \end{pmatrix} + X'_i \eta_i + \varepsilon_i, \quad (10)$$

where $\ln S_i = (\ln S_{i1}, \dots, \ln S_{i,T_i})'$, $X_i = (X_{i1}, \dots, X_{i,T_i})$, $\varepsilon_i = (\varepsilon_{i1}, \dots, \varepsilon_{i,T_i})'$ and

$$X_i^* = \begin{pmatrix} X_{i1} & & & \\ & X_{i2} & & \\ & & \ddots & \\ & & & X_{i,T_i} \end{pmatrix}. \quad (11)$$

After some manipulations Equation (10) becomes

$$\ln S_i - X'_i \eta_i = X_i^{*'} (Z'_i \otimes \mathbf{I}_{K+1}) \text{vec}(\gamma) + \varepsilon_i, \quad (12)$$

where $Z_i = (Z_{i1}, \dots, Z_{i,T_i})$. This equation can compactly be written as

$$W_i = V_i \text{vec}(\gamma) + \varepsilon_i, \quad (13)$$

where $\varepsilon_i \sim N(0, \sigma_i^2 \mathbf{I}_{T_i})$. From Equation (13) it is easy to derive that the full conditional posterior distribution of $\text{vec}(\gamma)$ is normal with mean

$$\left(\sum_{i=1}^I \frac{1}{\sigma_i^2} V'_i V_i \right)^{-1} \left(\sum_{i=1}^I \frac{1}{\sigma_i^2} V'_i W_i \right), \quad (14)$$

and variance

$$\left(\sum_{i=1}^I \frac{1}{\sigma_i^2} V'_i V_i \right)^{-1} \quad (15)$$

(see, for example, Zellner 1971, Chapter III).

Sampling of η_i

The relevant equations for sampling η_i for $i = 1, \dots, I$ are

$$\begin{aligned} \frac{1}{\sigma_i} [\ln S_i - X_i' \text{vec}(\gamma Z_i)] &= \frac{1}{\sigma_i} X_i' \eta_i + \frac{1}{\sigma_i} \varepsilon_i \\ 0 &= \Sigma_\eta^{-1/2} \eta_i + \nu_i \\ \frac{1}{\sigma_\xi} (F_i - \delta) &= \frac{1}{\sigma_\xi} (\mathbf{1}_M \otimes \theta') \eta_i + \frac{1}{\sigma_\xi} \xi_i, \end{aligned} \quad (16)$$

where $\nu_i \sim N(0, \mathbf{I}_{K+1})$, $\mathbf{1}_M$ is an $M \times 1$ vector of ones and $F_i = (\text{facings}_{i, T_1}, \dots, \text{facings}_{i, T_M})'$. The second line in Equation (16) represents the second layer of our model, and the third line represents the facings equation. We see that the full conditional posterior distribution of η_i conditional on γ , σ_i^2 , δ , θ , and σ_ξ^2 is normal. Denoting the first equation of (16) by $\tilde{Y}_i = \tilde{X}_i' \eta_i + \tilde{\varepsilon}_i$, the mean of this posterior distribution is

$$\left(\tilde{X}_i' \tilde{X}_i + \Sigma_\eta^{-1} + \frac{M}{\sigma_\xi^2} \theta \theta' \right)^{-1} \left(\tilde{X}_i' \tilde{Y}_i + \frac{1}{\sigma_\xi^2} (\mathbf{1}_M' \otimes \theta) (F_i - \delta) \right), \quad (17)$$

and the variance equals $(\tilde{X}_i' \tilde{X}_i + \Sigma_\eta^{-1} + (M/\sigma_\xi^2) \theta \theta')^{-1}$.

Sampling of σ_i^2

Conditional on the data and the other parameters, σ_i^2 has an inverted Gamma-2 distribution with scale parameter $\sum_{t=1}^{T_i} \varepsilon_{it}(\gamma, \eta_i)^2$, and degrees of freedom T_i . To sample σ_i^2 , we use that

$$\frac{\sum_{t=1}^{T_i} \varepsilon_{it}(\gamma, \eta_i)^2}{\sigma_i^2} \sim \chi^2(T_i), \quad (18)$$

where $\varepsilon_{i,t}$ is given in Equation (8).

Sampling of Σ_η

Conditional on the other parameters, the covariance matrix Σ_η can be sampled from an inverted Wishart distribution with scale parameter $\sum_{i=1}^I \eta_i \eta_i' + S$ and degrees of freedom $I + \lambda$.

Sampling of δ and θ

To derive the conditional posterior distribution of δ and θ , we collect the information on the facings of all items and all shelf allocation periods in one system

$$F = (\mathbf{1}, (\eta' \otimes \mathbf{1}_M)) \begin{pmatrix} \delta \\ \theta \end{pmatrix} + \xi, \quad (19)$$

where $F = (F_1', \dots, F_I)'$, η is the $(K+1) \times I$ matrix obtained by concatenating all random effects, that is, $\eta = (\eta_1, \dots, \eta_I)$, and $\xi = (\xi_1, \dots, \xi_I)'$, with $\xi_i = (\xi_{i1}, \dots, \xi_{iM})'$.

Again, it is easily seen that conditional on the data and other parameters, draws for $(\delta, \theta)'$ can be obtained from a normal distribution. Denoting $H = (\mathbf{1}, (\eta' \otimes \mathbf{1}_M))$, the mean of this distribution is $(H'H)^{-1} H'F$ and the variance is $\sigma_\xi^2 (H'H)^{-1}$.

Sampling of σ_ξ^2

The sampling of σ_ξ^2 is analogous to the sampling of σ_i^2 . We can use that, conditional on the other parameter and data,

$$\sum_{i=1}^I \sum_{m=1}^M \frac{\xi(\delta, \theta, \eta_i)^2}{\sigma_\xi^2} \sim \chi^2(MI), \quad (20)$$

where $\xi(\delta, \theta, \eta_i)$ is defined in (8).

References

- Blattberg, R. C., E. I. George. 1991. Shrinkage estimation of price and promotional elasticities: Seemingly unrelated equations. *J. Amer. Statist. Assoc.* **86** 304–315.
- Boatwright, P., R. E. McCulloch, P. E. Rossi. 1999. Account-level modeling for trade promotion: An application of a constrained parameter hierarchical model. *J. Amer. Statist. Assoc.* **94** 1063–1073.
- Borin, N., P. W. Farris. 1995. A sensitivity analysis of retailer shelf management models. *J. Retailing* **71** 153–171.
- Borin, N., P. W. Farris, J. R. Freeland. 1994. A model for determining retail product category assortment and shelf space allocation. *Decision Sci.* **25** 359–384.
- Brown, W., W. T. Tucker. 1961. The marketing center: Vanishing shelf space. *Atlanta Econom. Rev.* **11** 9–13.
- Bultez, A., P. Naert. 1988. SH.A.R.P.: Shelf allocation for retailers' profit. *Marketing Sci.* **7**(3) 211–231.
- Casella, G., E. I. George. 1992. Explaining the Gibbs sampler. *Amer. Statistician* **46** 167–174.
- Corstjens, M., P. Doyle. 1981. A model for optimizing retail space allocations. *Management Sci.* **27**(7) 822–833.
- Corstjens, M., P. Doyle. 1983. A dynamic model for strategically allocating retail space. *J. Oper. Res. Soc.* **34** 943–951.
- Cox, K. K. 1970. The effect of shelf space upon sales of branded products. *J. Marketing Res.* **7** 55–58.
- Curhan, R. 1972. The relationship between shelf space and unit sales. *J. Marketing Res.* **9** 406–412.
- Drèze, X., S. J. Hoch, M. E. Purk. 1994. Shelf management and space elasticity. *J. Retailing* **70** 301–326.
- Fok, D., Cs. Horváth, R. Paap, P. H. Franses. 2006. A hierarchical Bayes error correction model to explain dynamic effects of price changes. *J. Marketing Res.* **43** 443–461.
- Geman, S., D. Geman. 1984. Stochastic relaxations, Gibbs distributions, and the Bayesian restoration of images. *IEEE Trans. Pattern Anal. Machine Intelligence* **6** 721–741.
- Geylani, T., A. J. Dukes, K. Srinivasan. 2007. Strategic manufacturer response to a dominant retailer. *Marketing Sci.* **26**(2) 164–178.
- Hobert, J. P., G. Casella. 1996. The effect of improper priors on Gibbs sampling in hierarchical linear mixed models. *J. Amer. Statist. Assoc.* **91** 1461–1473.
- Johnson, D. S., C. R. Aragon, L. A. McGeoch, C. Schevon. 1989. Optimization by simulated annealing—An experimental evaluation. *Oper. Res.* **37**(6) 865–892.
- Kirkpatrick, S., C. Gelatt, M. Vecchi. 1983. Optimization by simulated annealing. *Science* **220** 671–680.
- Larson, J. S., E. Bradlow, P. Fader. 2005. An exploratory look at supermarket shopping paths. *Internat. J. Res. Marketing* **22** 395–414.
- Lim, A., B. Rodrigues, X. Zhang. 2004. Metaheuristics with local search techniques for retail shelf-space optimization. *Management Sci.* **50** 117–131.
- Luo, L., P. Kannan, B. T. Ratchford. 2007. New product development under channel acceptance. *Marketing Sci.* **26**(2) 149–163.
- Montgomery, A. L. 1997. Creating micro-marketing pricing strategies using supermarket scanner data. *Marketing Sci.* **16**(4) 315–337.
- Silva-Risso, J. M., R. E. Bucklin, D. G. Morrison. 1999. A decision support system for planning manufacturers' sales promotion calendars. *Marketing Sci.* **18**(3) 274–300.
- Sloot, L., D. Fok, P. Verhoef. 2006. The short- and long-term impact of an assortment reduction on category sales. *J. Marketing Res.* **43** 536–548.
- Tanner, M. A., W. H. Wong. 1987. The calculation of posterior distributions by data augmentation. *J. Amer. Statist. Assoc.* **82** 528–550.
- Tierney, L. 1994. Markov chains for exploring posterior distributions. *Ann. Statist.* **22** 1701–1762.

- van Dijk, A., H. van Heerde, P. Leeflang, D. Wittink. 2004. Similarity-based spatial methods to estimate shelf space elasticities. *Quant. Marketing Econom.* **2** 257–277.
- Wittink, D. R., M. J. Addona, W. J. Hawkes, J. C. Porter. 1988. SCAN-PRO: The estimation, validation and use of promotional effects based on scanner data. Technical report, Johnson Graduate School of Management, Cornell University, Ithaca, NY.
- Yang, M.-H. 2001. An efficient algorithm to allocate shelf space. *Eur. J. Oper. Res.* **131** 107–118.
- Yang, M.-H., W.-C. Chen. 1999. A study on shelf space allocation and management. *Internat. J. Production Econom.* **60–61** 309–317.
- Zellner, A. 1971. *An Introduction to Bayesian Inference in Econometrics*. John Wiley, New York.