Management Science

Publication details, including instructions for authors and subscription information:
http://pubsonline.informs.org

A Demand Estimation Procedure for Retail Assortment Optimization with Results from Implementations

Marshall Fisher, Ramnath Vaidyanathan

MANAGEMENT SCIENCE
Vol. 60, No. 10, October 2014, pp. 2401–2415
ISSN 0025-1909 (print) | ISSN 1526-5501 (online)

http://dx.doi.org/10.1287/mnsc.2014.1904

A Demand Estimation Procedure for Retail Assortment Optimization with Results from Implementations

Marshall Fisher
Operations and Information Management Department, The Wharton School, University of Pennsylvania, Philadelphia, Pennsylvania 19104, fisher@wharton.upenn.edu

Ramnath Vaidyanathan
Desautels Faculty of Management, McGill University, Montreal, Quebec H3A 1G5, Canada, ramnath.vaidyanathan@mcgill.ca

We consider the problem of choosing, from a set of N potential stock-keeping units (SKUs) in a retail category, K SKUs to be carried at each store to maximize revenue or profit. Assortments can vary by store, subject to a maximum number of different assortments. We view a SKU as a set of attribute levels and also model possible substitutions when a customer's first choice is not in the assortment. We apply maximum likelihood estimation to sales history of the SKUs currently carried by the retailer to estimate the demand for attribute levels and substitution probabilities, and from this, the demand for any potential SKU, including those not currently carried by the retailer. We specify several alternative heuristics for choosing SKUs to be carried in an assortment. We apply this approach to optimize assortments for three real examples: snack cakes, tires, and automotive appearance chemicals. A portion of our recommendations for tires and appearance chemicals were implemented and produced sales increases of 5.8% and 3.6%, respectively, which are significant improvements relative to typical retailer annual comparable store revenue increases. We also forecast sales shares of 1, 11, and 25 new SKUs for the snack cake, tire, and automotive appearance chemical applications, respectively, with mean absolute percentage errors (MAPEs) of 16.2%, 19.1%, and 28.7%, which compares favorably to the 30.7% MAPE for chain sales of two new SKUs reported by Fader and Hardie (1996).

Keywords: assortment planning; retailer operations; statistics; estimation
History: Received June 16, 2009; accepted November 13, 2013, by Yossi Aviv, operations management. Published online in *Articles in Advance* July 10, 2014.

1. Introduction

A retailer's assortment is the set of products they carry in each store at each point in time. Retailers periodically revise the assortment for each category, deleting some products and adding others to take account of changes in customer demand over time as well as new products introduced by suppliers. This periodic assortment reset seeks to choose a set of stock-keeping units (SKUs) to carry in the new assortment to maximize revenue or profit over a future planning horizon, subject to a shelf space constraint, which can often be expressed as an upper bound on the number of SKUs carried.

A customer whose most preferred product is not in a retailer's assortment may elect to buy nothing or to purchase another product sufficiently similar to their most preferred product. These substitution probabilities need to be estimated, and the sales of a SKU to customers who most preferred that SKU must be distinguished from sales to customers who preferred a different SKU but substituted when they did not find their preferred SKU in the assortment.

Most retailers use the same assortment in all stores, except for eliminating some SKUs in smaller stores. Recently, however, localizing assortments by store or store cluster has become a high priority for many retailers. For example, Zimmerman (2006, 2008), O'Connell (2008), and McGregor (2008) describe recent efforts by Wal-Mart, Home Depot, Macy's, and Best Buy to vary the assortment they carry at each store to account for local tastes. Despite the flurry of interest in localization reported in the business press and that we have encountered in our interaction with retailers, there have been no studies to document the level of benefits from localization or to provide tools to help a retailer determine the right degree of localization.[1]

SKU deletion decisions are relatively easy, since sales data indicate the popularity of existing SKUs,

[1] This discussion is based on Fisher and Raman (2010) and conversations with several retail executives, including Paul Beswick, partner and head of Oliver Wyman North American Retail Practice; Robert DiRomualdo, former chief executive officer of Borders Group; Kevin Freeland, chief operating officer of Advance Auto; Matthew Hamory, principal of Oliver Wyman North American Retail Practice; Herbert Kleinberger, principal of ARC Consulting; Chris Morrison, senior vice president of sales, Americas for Tradestone; Robert Price, chief marketing officer of CVS; and Cheryl Sullivan, vice president of product management of Revionics, Inc.

but retailers lack (with the partial exception of grocery, where companies like Nielsen and IRI provide industry retail sales data for a wide range of products) analytic tools to forecast the sales of SKUs that might be added to the assortment, to estimate the rate of customer substitution, and to intelligently localize assortments by store. The goal of this paper is to provide these tools.

Our approach follows the marketing literature in viewing a SKU as a set of attribute levels and assuming that a given customer has a preferred set of attribute levels. We use prior sales to estimate the market share in each store of each attribute level and forecast the demand share for a SKU as the product of the demand shares for the attribute levels of that SKU. We assume that if a customer does not find their ideal product in the assortment, they buy the product in the assortment closest to their ideal with some probability, and we estimate these substitution probabilities. We apply various heuristics to these demand and substitution estimates to determine optimized assortments.

In the extreme, a retailer might carry a unique assortment in each store, but most retailers claim that this is administratively too complicated. In particular, retailers develop a diagram called a planogram showing how all products should be displayed in a store, a process that is labor intensive. A planogram would need to be developed for each assortment, which means the administrative cost of each assortment is high. Our process can control the degree of localization by limiting the number of different assortments to any level between a single assortment for the chain and a unique assortment for each store.

We applied this approach to the snack cakes category at a regional convenience store chain, the tire assortment at a national tire retailer, and the appearance chemicals category of a major auto aftermarket parts retailer. The tire and auto parts retailers implemented portions of our recommended assortment changes and obtained revenue increases of 5.8% and 3.6%, respectively, significant improvements given traditional comparable store annual increases in these segments.

Our approach best fits situations where demand is changing over time, but gradually enough that history is useful; where the individual consumer data common in grocery are not available; where some products are better substitutes for a given product than others; and when inventory decisions are unrelated to assortment decisions, because the retailer carries a fixed, often small amount of inventory for each SKU. This was the case for all of our applications.

Despite the enormous economic importance of assortment planning, we are aware of only two papers, Chong et al. (2001) and Kök and Fisher (2007), that formulate a decision support process for assortment planning and test the process on real data. This paper extends these papers, and hence the existing literature, in four ways.

1. We provide a thorough treatment of assortment localization, allowing a constraint on the number of different assortments and measuring how the level of localization impacts revenue. Chong et al. (2001) do not deal with localization. Kök and Fisher (2007) allow a unique assortment for each store, but do not provide a constraint on the number of assortments or assess how localization impacts revenue.

2. We forecast the demand for new SKUs that have not been carried before in any store, based on past sales of products currently carried. Chong et al. (2001) need consumer-level transaction data over multiple shopping trips for new SKU forecasting, and these detailed data are not available in most nongrocery applications. Kök and Fisher (2007) forecast how a SKU carried in some stores will sell in other stores in which it is not carried, but do not provide a way to forecast demand for a completely new SKU carried in no stores.

3. We introduce a new demand model that fits a situation in which some products are better substitutes for a given product than others. This was a real feature of our applications; for example, the natural substitutes for a 14-inch tire are other 14-inch tires, not 15-inch tires.

4. Two of the retailers in our study implemented a portion of our recommended assortment changes, and we estimate that the revenue increase from these changes is larger than the annual same store revenue increases typically seen in these industries. We believe this is the first implementation validation of an analytic approach to assortment planning.

2. Literature Review

We review here the prior research on consumer choice models, demand estimation, and assortment optimization that is most related to this paper. See Kök et al. (2009) for a more extensive review.

Consumer choice models constitute the fundamental platform for assortment planning. Utility based models, such as the multinomial logit (MNL) model (Guadagni and Little 1983), assume that every customer associates a utility U_i with each SKU $i \in N$. In addition, there is a no-purchase option denoted $i = 0$, with associated utility U_0. When offered an assortment A, every customer chooses the option giving him the highest utility in $A \cup \{0\}$. The market share for each SKU $i \in A$ can then be evaluated once we know the distribution of utilities across the consumer population.

In the locational choice model (Hotelling 1929, Lancaster 1975), every SKU $i \in N$ is represented as a bundles of attribute levels. Each consumer has an ideal set of attribute levels and buys the product with these attributes if it is in the assortment. Otherwise, he buys the product in the assortment closest to his ideal product if it is close enough to offer positive utility, or else he elects not to purchase.

In the exogenous demand model, every consumer is assumed to have a favorite product i, and f_i is the share of consumers whose favorite product is i. A consumer whose favorite product is i buys it if $i \in A$; if $i \notin A$, they substitute to SKU $j \in A$ with probability α_{ij}. Under these assumptions, the market share of each SKU $i \in A$ is given by $q_i(A) = f_i + \sum_{j \notin A} f_j \alpha_{ji}$.

All three demand models assume that customers have a favorite product that they buy if it is in the assortment, or they may substitute a different product if their favorite product is not in the assortment. Where the models differ most is in their assumptions about substitution behavior. The exogenous demand model allows any substitution structure, but has many parameters that are difficult to estimate in practice. The MNL model assumes that the demand for a missing product that is not lost transfers to other products in the assortment in proportion to their popularity. By contrast, the locational choice model assumes that a given product may be more like some products than others and that substitution demand transfers to the product in the assortment that is most similar to a customer's preferred product.

Demand estimation has been studied by Fader and Hardie (1996), who use maximum likelihood estimation (MLE) to estimate the parameters of an MNL model from sales transaction data under the assumption that product utility is the sum of utilities of the product attributes. Anupindi et al. (1998), Chong et al. (2001), Bell et al. (2005), Kök and Fisher (2007), and Vulcano et al. (2012) have also studied demand estimation.

Assortment optimization research has been based on both stylized models intended to provide insight into structural properties of optimal assortments and decision support models intended to guide a manager planning retail assortments.

Stylized model research began with van Ryzin and Mahajan (1999), who show that an optimal assortment under an MNL consumer choice model consists of a certain number of the highest utility products.

Gaur and Honhon (2006) show that for a locational choice model, the products in the optimal assortment are located far from each other in the attribute space, with no substitution between products. This implies that the most popular product may not be carried in the optimal assortment, contrasting the results of van Ryzin and Mahajan (1999).

Decision support research began with Green and Krieger (1985), who formulate the problem of which out of a set of potential products a firm should select to maximize (1) consumer welfare or (2) firm profits. Belloni et al. (2008) compare the performance of different heuristics for product line design and find that the greedy and the greedy-interchange heuristics perform extremely well.

Smith and Agrawal (2000) use an exogenous demand model and an integer programming formulation of assortment planning. They solve a number of small problems by complete enumeration to demonstrate how assortment and stocking decisions depend on the nature of assumed substitution behavior, and also propose a heuristic to solve larger problems.

Chong et al. (2001) develop an assortment modeling framework using consumer-level transaction data over multiple grocery shopping trips to estimate the parameters of the model and use a local improvement heuristic to suggest an alternative assortment with higher revenue. Kök and Fisher (2007) use an exogenous demand model to study a joint assortment selection and inventory planning problem at a large Dutch grocery retailer.

3. Problem Formulation and Demand Model

We seek optimal assortments for a retail category over a specified future planning horizon. We assume that the goal is to maximize revenue, since this was the concern in our three applications. It is straightforward to adapt our approach to maximizing other functions, such as unit sales or gross profit. We define $N = \{1, 2, \ldots, n\}$ to be the index set of all possible SKUs a retailer could carry in this category, and $M = \{1, 2, \ldots, m\}$ to be the index set of all stores. Let K be the maximum number of SKUs per assortment, and let L be the maximum number of different assortments. Finally, D_i^s is the number of customers at store s for whom SKU i is their most preferred product, and p_i is the price of SKU $i \in N$.

For expositional simplicity, K does not vary by store, but it would be easy to modify our process to allow K to vary by store, and in fact we do this in our computational work. We choose L to lie between 1 and m to trade off the greater revenue that comes with larger L against the administrative simplicity that comes with smaller L. As will be seen, our solution approach makes it easy to solve this problem for all possible values of K and L, thus providing the retailer with rich sensitivity analysis to guide their choice of these parameters.

We define an assortment to be a set $S \subseteq N$ with $|S| \leq K$ and let l_s denote the index of the assortment assigned to store s. A solution to the assortment optimization problem is defined by a portfolio of assortments S_l, $l = 1, 2, \ldots, L$ and $l_s \in \{1, 2, \ldots, L\}$, for all $s \in M$.

Our demand model assumes that a consumer shopping in this category in store s has a most preferred

SKU $i \in N$, but might be willing to substitute to other SKUs if $i \notin S_{l_s}$. We view a SKU as a collection of attribute levels, use historical sales data to estimate the demand share of each attribute level, and finally estimate the demand share of any SKU as the product of the demand shares of its attribute levels.

We introduce some notation to formalize this approach. Let A be the number of attributes, let a be the attribute type index, and let N_a be the number of levels of attribute $a \in \{1, 2, \ldots, A\}$. Furthermore, let f^s_{au} denote the fraction of customers at store s who prefer level u of the attribute a, where $u = 1, 2, \ldots, N_a$ and $a = 1, 2, \ldots, A$. Finally, let π^s_{auv} be the probability that a customer at store s whose first choice on attribute a is u is willing to substitute to v, defined for all a, u, and v.

By definition, $\pi^s_{avv} = 1$. Moreover, π^s_{auv} can be 0 if attribute level v is not a feasible substitute for u.

The fraction of customers who most prefer SKU i with attribute levels $i_1, i_2, \ldots, i_A$ is defined to be $f^s_i = \Pi^{a=A}_{a=1} f^s_{ai_a}$. If a customer's most preferred SKU i with attributes $i_1, i_2, \ldots, i_A$ is not in the assortment, they are willing to substitute to SKU j with attributes $j_1, j_2, \ldots, j_A$ with probability $\pi^s_{ij} = \Pi^{a=A}_{a=1} \pi^s_{ai_a j_a}$. If a customer with most preferred SKU i finds $i \in S_{a(s)}$ when they shop the store, then we assume they buy it. Otherwise, they buy the best substitute for i in S, defined to be $j(i, S) = \arg\max_{j \in S} \Pi^A_{a=1} \pi^s_{ai_a j_a}$ with probability $\pi^s_{ij(i,S)}$.

In using store sales data to estimate the parameters of our model, we will estimate D^s, the total category unit demand in store s, if all possible SKUs were offered, and then set $D^s_i = f^s_i D^s$.

The revenue earned by store s using assortment S_{l_s} can then be written as

$$R_s(S_{l_s}) = \left(\sum_{i \in S_{l_s}} p_i D^s_i + \sum_{i \notin S_{l_s}} D^s_i \pi^s_{ij(i, S_{l_s})} p_{j(i, S_{l_s})} \right).$$

The first term in this expression is the revenue from customers whose most preferred SKU is contained in the assortment, and the second term is the expected substitution revenue from customers whose most preferred SKU was not in the assortment.

The assortment optimization problem is to choose S_l, $|S_l| \leq K$, $l = 1, 2, \ldots, L$ and l_s for all $s \in M$ to maximize $\sum_{s \in M} R_s(S_{l_s})$.

We allow substitution probabilities to vary by store because we found in our applications that they did in fact vary by store. For example, in the tire application, the willingness of a consumer to substitute to a higher priced product varied by store and was correlated with median income in the zip code in which the store is located.

Our demand model implies that a consumer's preferences for the various attributes are independent, which may not be true. For example, a college student shopping for twin-size bed sheets might have a different color preference than a suburban homemaker shopping for queen-size sheets, so the color and size attributes for sheets would interact.

Our defense of this assumption is threefold: (1) All prior publications we are aware of that use attributes in demand estimation make a similar assumption. For example, Fader and Hardie (1996) assume that the utility for a SKU is a linear function of its attributes and then use this utility in a multinomial logit model to determine SKU demand shares. They state that "[i]n both the marketing and economics literature, it is common to assume an additive utility function" (p. 444) and note that this implies no interaction between attributes. (2) In our applications, we check the accuracy of this approximation by comparing demand estimates with actual sales for the SKUs currently carried and find that forecasts based on this model are accurate compared to previously published research. (3) If there is significant interaction between attributes, we demonstrate ways to modify our demand model to take this into account. In the snack cake application, the attributes are flavor, package size (single serve or family size), and brand. Package size and brand interact since one brand is stronger in single serve and another in family size. We deal with this by combining brand and size into a new attribute, brand-size. In the tire application, the attributes are size, brand, and mileage warranty. Brand and mileage warranty interact because a given brand does not offer all warranty levels, and so we combine brand and warranty level to create the attribute brand-warranty.

In the tire application, there is also an interaction between size and brand-warranty. A tire with a given size attribute level fits a defined set of car models of a certain age and value. The brand-warranty levels correspond to different price points and quality and might interact with the age and value of the cars a size tire fits. We show how to deal with this by partitioning the sizes into a finite number of homogeneous segments (which are latent) and allowing the brand-warranty shares to be conditional on segment membership (Fader and Hardie 1996, Kamakura and Russell 1989).

4. Analysis

We describe our methods for estimating model parameters and choosing assortments.

4.1. Estimating Demand and Substitution Probabilities

We use maximum likelihood estimation to estimate demand and substitution probabilities. Our primary input for estimation is store-SKU sales of products currently carried by the retailer during a prior history period. Parameters are estimated at the store level, but for expositional simplicity we will drop the store superscript in the discussion that follows. We describe here a generic, broadly applicable approach; in §5, we will exploit special structure of the applications to refine this approach. Let S denote the assortment carried in a particular store, and let x_i denote the sales of SKU $i \in S$ during a history period.

We can write the probability $F_j(S)$ that a customer purchases $j \in S$ as $F_j(S) = f_j + \sum_{i \notin S,\, j=j(i,S)} f_i \pi_{ij}$. Let $F(S) = \sum_{j \in S} F_j(S)$ denote the probability that a customer shopping in this category makes any purchase from assortment S. Then, assuming that each consumer purchase is an independent random draw, the likelihood of observing sales data $x = \{x_i\}_{i \in S}$ is given by $LH(f, \pi) = C \prod_{j \in S} [F_j(S)/F(S)]^{x_j}$, where the proportionality constant is $C = (\sum x_{j \in S})!/\prod_{j \in S} x_j!$. The maximum likelihood estimates for the parameters (f, π) can be obtained by maximizing the log-likelihood function

$$LLH(f, \pi) = \sum_{j \in S} x_j \log F_j(S) - \left(\sum_{j \in S} x_j\right) \log F(S) \quad (1)$$

subject to the constraints

$$\sum_{u=1}^{N_a} f_{au} = 1, \quad a = 1, 2, \ldots, A, \quad (2)$$

$$f_i = \prod_{a=1}^{A} f_{ai_a}, \quad i = 1, 2, \ldots, n, \quad (3)$$

$$\pi_{ij} = \prod_{a=1}^{A} \pi_{ai_a j_a} \quad i = 1, 2, \ldots, n \text{ and } j = 1, 2, \ldots, n, \quad (4)$$

$$f_{au}, \pi_{auv} \in [0, 1], \quad \forall\, a,\, u, \text{ and } v. \quad (5)$$

Given the complex nature of the log-likelihood function, it is not possible to derive analytical results. Hence, we resort to numerical optimization methods based on gradients after transforming the problem into an unconstrained optimization problem by reparametrizing f_{au} and π_{auv} in terms of $\hat{f}_{au}$ and $\hat{\pi}_{auv}$, where

$$f_{au} = \frac{\exp(\hat{f}_{au})}{\sum_{u=1}^{N_a} \exp(\hat{f}_{au})}, \quad u = 1, 2, \ldots, N_a - 1 \text{ and } \hat{f}_{aN_a} = 1, \quad (6)$$

$$\pi_{auv} = \frac{\exp(\hat{\pi}_{auv})}{1 + \exp(\hat{\pi}_{auv})}, \quad \forall\, a, u, \text{ and } v. \quad (7)$$

We found examples showing that the log-likelihood function may not be concave, which implies that numerical optimization methods may not converge to a global maximum. We handle this issue by running the optimization algorithm from several randomly generated starting points. This does not guarantee global convergence, but lowers the chances of the algorithm getting stuck at a local maximum. Mahajan and van Ryzin (2001) use a similar approach to compute the optimal inventory levels using the sample path gradient algorithm. Once we obtain MLE estimates for demand shares and substitution probabilities, as noted in the previous section, we estimate total demand for the product category as $D = \sum_{i \in S} x_i / F(S)$, and D_i as $f_i D$.

4.2. Estimating Prices for New SKUs

We set prices on SKUs not currently carried by the retailer to be consistent with their current pricing policy. We assume that prices on existing SKUs were set in relationship to the value of the SKU to a consumer and that consumer value is related to attribute levels. Hence, we regressed the log of price on attribute levels to obtain the pricing equation:

$$\log(p_i) = \alpha_0 + \sum_{a=1}^{A} \sum_{u=1}^{N_a - 1} \beta_{au} z_{iau}, \quad i = 1, 2, \ldots, n, \quad (8)$$

where z_{iau} is a dummy variable taking the value one if SKU i has level u of attribute a, and zero otherwise.[2]

This is a hedonic pricing equation and has been used extensively in economics (Rosen 1974).

4.3. Heuristics for Choosing Assortments

Absent substitution, the assortment problem could be optimally solved by a greedy algorithm that chooses SKUs in decreasing order of their revenue contribution. But substitution makes the objective function nonlinear, because the contribution of a SKU depends on its substitution demand, which depends on the other SKUs in the assortment. As a result, the assortment problem is complex to solve optimally. Hence, we define greedy and interchange heuristics for choosing the assortments S_l, $l = 1, 2, \ldots, L$ and the specification l_s, $s \in M$ of the assortment assigned to store s.

For assortment planning, Kök and Fisher (2007) use a greedy heuristic and Chong et al. (2001) an interchange heuristic. For the product line design problem, Green and Krieger (1985) use greedy and interchange heuristics. Belloni et al. (2008) find that for the product line design problem, greedy and interchange heuristics together find 98.5% of optimal profits on average for randomly generated problems, and 99.9% for real problems.

The greedy heuristic for finding a single assortment for a store or set of stores adds SKUs in decreasing order of revenue contribution until K SKUs have been added. The interchange heuristic starts with a given assortment and tests whether interchanging a SKU that is not in the assortment with a SKU in the assortment would increase revenue. Any revenue increasing interchanges are made as they are discovered. The process continues

[2] The summation is only over $N_a - 1$ levels of attribute a, since the reference levels for all attributes are accounted for by the intercept term α_0.

until a full pass over all possible interchanges discovers no revenue increasing interchanges. We apply the interchange heuristic both starting with the greedy assortment and starting with random assortments.

To find a portfolio of L assortments and assignments of stores to assortments, we have two alternative heuristics, a forward and reverse greedy. In the forward greedy heuristic, we first apply the greedy and interchange heuristics to generate a single assortment for the chain. If $L > 1$, we apply greedy and interchange heuristics to generate an assortment for each store and add as a second assortment the store assortment that maximizes the revenue increase with stores optimally assigned to each assortment. This process continues until L assortments have been chosen.

In the reverse greedy heuristic, we first apply greedy and interchange heuristics to generate an assortment for each store. If $L < m$, we identify the single store assortment to delete that would minimize the revenue loss. We calculate the revenue loss from deleting an assortment by reassigning stores to their revenue maximizing assortment in the reduced set of assortments and calculating the loss in revenue due to the reassignment. At any point in the algorithm, we have a portfolio of l assortments and l store clusters defined by the assignment of stores to assortments. As long as $l > L$, we delete one assortment from this portfolio that leads to the least loss in revenue and reassign stores to the reduced set of assortments.

To evaluate the performance of the heuristics, we proved that if we use estimated rather than actual prices for existing SKUs, the optimal solution in our applications has a special structure that allows us to find optimal assortments. We used this result to create optimal assortments for the 140 stores in the snack cake application and the 574 stores in the tire application, then applied the greedy heuristic and found it generated solutions that were 97.2% and 98.5% of the optimal, respectively. We also tested the interchange heuristic, starting both with the greedy assortment and random assortments, but in no case found solutions that improved on the greedy assortment.

5. Applications

We worked with three retailers to apply our methodology, analyzing one product category for each of them: snack cakes for a regional convenience chain, tires for a national tire retailer, and appearance chemicals for a major auto aftermarket parts retailer. Results for the three applications are summarized in Table 1. The entries in this table are described below.

5.1. Description of Applications and SKU Attributes

The regional convenience chain offered snack cakes in 60 flavors, two brands, $\{B_1, B_2\}$, and several different package sizes in 140 stores. We restricted our analysis to the top 23 flavors that accounted for 95% of revenue. Although there were several different package sizes, what mattered to consumers was whether the size was single serve or family size. So we grouped sizes into *single serve* (*Si*) and *family size* (*Fa*). Furthermore, because the retailer advised us that brand shares and willingness to substitute varied by size, we combined brand and size to obtain a single attribute called *brand-size*, indexed 1 to 4 for B_1Si, B_2Si, B_1Fa, and B_2Fa in order.

Table 1 Summary of Key Values for the Three Applications

	Application		
	Snack cakes	Tires	Appearance chemicals
Category parameters			
n—Possible SKU count	92	384	154
K—Average number of SKUs in an assortment	40	105	130
m—Store count	140	574	3,236
Demand forecasts (%)			
Fit MAD (chain, store)	6.2, 16.4	4.5, 13.6	8.5, 19.3
Validation MAD (chain, store)	25.8, 40.1	21.1, 38.2	
Price regression R^2	85.5	96.3	
Estimated revenue increase (based on fit sample)			
Chain optimal assortment (%)	29.2	30.1	11.8
Store optimal assortment (%)	41.4	35.9	14.2
Implementation (%)		5.8	3.6
Number of new SKUs forecasted	1	11	25
New SKU forecast MAPE (%)	16.2	19.1	28.7
Computation time (average seconds per store)			
Estimation	7.2	13.3	5
Greedy	2.2	6.5	1.4

Thus, there were 23 *flavor* attribute levels, 4 *brand-size* attribute levels, and 92 possible SKUs, of which 52 were being offered by the retailer in at least one store. The number of SKUs offered across stores varied between 24 and 52, and averaged 40.3. An internal market research study on the industry commissioned by the retailer showed that flavor was the most important attribute for a consumer. So, we assumed that the probability of substituting across flavors could be set to 0. The retailer also believed that there is negligible substitution between sizes S and F, so this substitution was assumed to be 0. The substitute for a particular brand-size is the other brand in the same size. We define $j(i)$ to be the brand-size that a customer might substitute to if i is not in the assortment, and set $j(1)=2$, $j(2)=1$, $j(3)=4$, and $j(4)=3$. We need to estimate the 23 flavor shares f_{1v}, 4 brand-size shares f_{2b}, and 4 substitution probability parameters π_{12}, π_{21}, π_{34}, and π_{43}.

Attributes for the tire study were brand, mileage warranty, and size. The retailer offered several nationally advertised brands that they believed were equivalent to the consumer, which we denote as National (N) and treat as one brand. They also offered three house brands of decreasing quality, which we denote as House 1 (H1), House 2 (H2), and House 3 (H3), where H1 is the highest quality and most expensive house brand. There were a large number of distinct mileage warranties offered, but some of these varied only slightly and hence were believed by the retailer to be equivalent to consumers. Therefore, we aggregated the mileage warranties into three levels of low (15,000–40,000 miles), medium (40,001–60,000 miles), and high ($>60{,}000$ miles), denoted L, M, and H, respectively. We combined brand and warranty into a single attribute to account for interaction between these attributes (e.g., national brands were always offered only with a high or medium warranty, whereas H3 was offered only with a low warranty) to identify the following six brand-warranty combinations: NH, NM, H1H, H2H, H2M, and H3L. Sixty-four distinct tire sizes were offered, resulting in $64\times 6=384$ possible tire SKUs that could be offered. This retailer carried 122 of these 384 possible SKUs in at least one of their stores. The number of SKUs offered across stores varied between 93 and 117, and averaged 105.2. The assortment offered also varied slightly across the stores, indicating some localization.

We were advised by the retailer that customers do not substitute across sizes. Table 2 was provided by the vice president of the tire category and depicts the qualitative likelihood of substitution across brand-warranty levels. We let $\{\pi_S, \pi_L, \pi_M\}$ denote the substitution probabilities *somewhat likely*, *likely*, and *most likely*.

The auto aftermarket parts retailer examines performance of each of their product categories once a year on a rotating schedule and considers changes in the assortment. We were asked in early May 2009 to apply our methodology for the annual assortment reset for the appearance chemicals category, a category comprised of liquids and pastes for washing, waxing, polishing, protecting, etc., all surfaces of an auto, including the body, tires, wheels, windshield and other glass, and various interior surfaces. The retailer used a market research firm, NPD, that had assigned to each SKU in the appearance category the attributes (1) segment (defined by the surface of the car treated and what is done to that surface), (2) brand (one of nine brands), and (3) quality level (one of three levels, denoted good, better, or best, where "good" is the lowest quality and "best" is the highest). We appended package size, denoted small (S) or large (L), to the segment attribute to create 45 segment-size attribute levels. We combined brand and quality to create a second attribute, brand-quality. Because some brands did not offer all quality levels, there were 17, not 27, brand-quality combinations. In some cases there were two package sizes that differed slightly and were classified as S or L, so that two SKUs occupied the same cell of the attribute matrix. Consequently, the 160 SKUs currently offered corresponded to 130 cells of the 45×17 matrix of possible attribute levels. Of the $45\times 17-130=635$ attribute combinations not carried by the retailer, only 24 were available in the market. No substitution parameters were used in the model for the following reason. If a brand-quality level was not offered for a product, it seemed likely that some of the demand for that brand-quality would transfer to several other brands. To estimate this effect, we would have needed instances of different stores with varying numbers of brand-quality levels offered for the same product, and these data were not available.

For each application, assortments varied somewhat across stores, although the retailers told us this was due to varying store sizes and was not a result of systematic efforts to localize assortments to store demand patterns. Typically, a small store would carry a subset of a large store assortment.

Table 2 **Tires: Management's Estimate of the Most Likely Substitution Probabilities**

	To					
	NH	NM	H1H	H2H	H2M	H3L
From						
NH	1	*S*	*S*	*S*	0	0
NM	*L*	1	*S*	*S*	0	0
H1H	0	0	1	*L*	*S*	0
H2H	0	0	*S*	1	*S*	0
H2M	0	0	*S*	*L*	1	0
H3L	0	0	0	0	*M*	1

Note. *S*, somewhat likely; *L*, likely; *M*, most likely.

Given this, we developed a measure of assortment commonality that would indicate the extent of localization for the assortments that existed at the start of our project. For any pair of stores A and B, the overlap in assortment was computed as the ratio of the number of common SKUs to the minimum number of SKUs across the two stores. Note that this metric equals 1 if the assortment of one store is identical to or a subset of the other, indicating no localization. The average for all store pairs was 0.971 for cakes, 0.94 for tires, and 0.964 for appearance chemicals, indicating that the extent of localization was minimal.

5.2. Estimation

We used sales data for up to three time periods, labeled fit, validation, and implementation samples. For snack cakes, the fit and validation periods were July–December 2005 and July–December 2007, respectively. For tires, the fit and validation periods were July–December 2004 and January–June 2005, respectively. Results were implemented for tires and the impact on sales tracked for July–December 2005, the implementation sample. For appearance chemicals, the fit period was May 2008–April 2009. The results obtained in July 2009 were so compelling that the retailer decided to implement them in early January 2010 without validation. To evaluate the impact of implementation, we used store-SKU sales for the 27-week period January 23–July 8, 2010.

We applied the estimation procedure described in §4.1 to store-SKU sales data for the fit sample to estimate model parameters at each store for the three applications. This process worked except for 255 stores of the tire retailer where there was insufficient data to determine all six brand-warranty shares because there was no size in which brand-warranties H2M and H3L were both offered, making it impossible to identify the split of demand between H2M and H3L. Of these 255 stores, there were 52 stores where we could also not identify the share of H2M. We therefore regressed the H3L and H2M shares against median income for the 319 stores at which we could estimate all the parameters and used the regression estimate for these shares as needed. We then used MLE to estimate the remaining parameters.

Prices of potential new snack cake and tire SKUs were estimated using a hedonic regression as described in §4.2, with the R^2 values shown in Table 1. We multiplied the estimated prices by scale factors of 1.02 and 1.05, respectively, to equalize the estimated revenue at the chain level to the actual revenue, which will facilitate comparison of new optimized assortments with current revenue. The retailer of appearance chemicals sold branded products, so the prices of potential new SKUs were known.

5.3. Validation

We measured the overall estimation error across all stores, by computing the sales-weighted mean absolute deviation (MAD) of estimated store-SKU sales shares from actual store-SKU sales shares, as given by

$$\frac{\sum_{s\in M}\sum_{j\in S^s}|(x_j^s/\sum_j x_j^s)-F_j^s(S^s)|x_j^s}{\sum_{s\in M}\sum_{j\in S^s}(x_j^s/\sum_j x_j^s)},$$

where S^s is the set of SKUs in the assortment for store s. We measure MAD in terms of sales shares because unit sales are significantly influenced by overall growth or shrinkage in the category, whereas sales shares are not. Moreover, our assortment choices are determined solely by sales shares, so these are the parameters important to our analysis.

We validated our results for the snack cake and tire applications using store-SKU sales for the validation periods defined above.[3] We used the fit estimated parameters to compute the share of sales for each SKU for the validation period, at the store and chain levels, and compared it with the actual sales shares to obtain the MADs given in Table 1.

One way to deal with interaction between attributes is a latent class analysis. For tires, we might assume that there are several homogeneous segments of sizes, and brand-warranty shares vary across segments, but are not the same within each segment. To explore the feasibility of this approach, we estimated such a model for a subset of 20 stores, allowing for two size segments. We found that the sales-weighted MAD of sales share estimates was 15.3% compared to 19.6% for the case of one size segment.

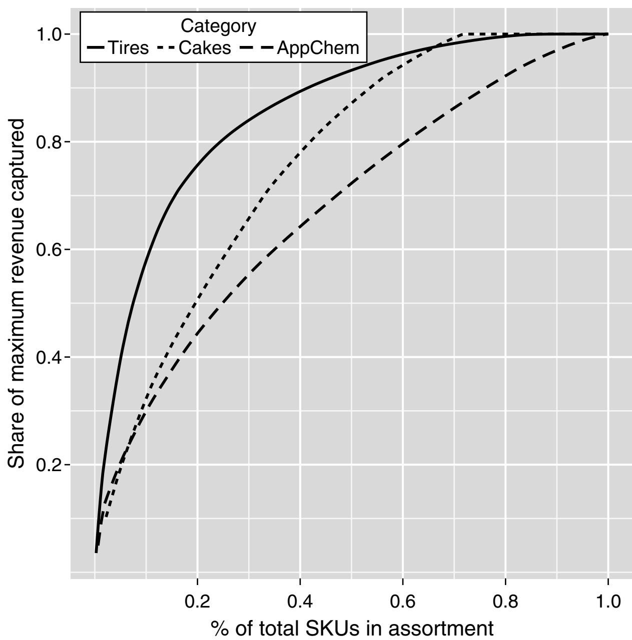
5.4. Optimization

We next applied the greedy heuristic described in §4.3 to compute optimized assortments at the chain and store levels, varying the number of SKUs in the assortment from 1 to n. Figure 1 shows the percentage of the maximum possible revenue if all SKUs were offered captured by the chain level assortment, as a function of K as a percentage of n. Note that maximizing revenue for a given value of K is equivalent to maximizing this percentage of maximum revenue captured.

For cakes and tires, the potential SKU set was all combinations of attributes. The appearance chemical application sold branded products available in the market place, and hence there was a list of available products that we used in the optimization. As noted previously, this comprised SKUs currently offered and 24 potential new SKUs. Interestingly, while we were waiting to receive this list from the category manager, we did an analysis using all combinations of

[3] For the snack cakes validation, we had data for only 54 of the 140 stores in the chain.

Figure 1 Revenue vs. Percentage of Maximum Possible SKUs in the Assortment



attributes, which recommended adding a wax product for a leading brand. The category manager pointed out that the product did not exist in the market place, but then was shocked when the brand sales representative met with her several days later and told her they were planning to introduce exactly this product. So for retailers of branded products, it makes sense to analyze two cases, one with SKUs that currently exist in the market place for assortment update and one with a SKU set corresponding to all attribute combinations, to identify new product opportunities for suppliers and private label opportunities for the retailer.

Table 1 summarizes computation times for estimation and greedy procedures on an Intel Core 2 Duo, 2GHz processor.

To quantify the potential improvement in revenue, we compared the assortments we generated for $L=1$ and $L=m$ to the current assortment. SKU count varied somewhat by store, so we compared the current assortment for store s with SKU count K_s to the greedy assortment for $K=K_s$.

5.5. Findings

Tables 3–5 give results of parameter estimation for the snack cake and tire applications. Tire substitution

Table 3 Snack Cakes: Demand Share Estimates (Averaged Across Stores)

Brand size	Sales share (%)	Demand share (%)
B_1Si	67	61
B_2Si	24	27
B_1Fa	4	6
B_2Fa	5	6

Table 4 Snack Cakes: Substitution Probability Estimates (Averaged Across Stores)

	B_1Si	B_2Si	B_1Fa	B_2Fa
B_1Si	1	18%	0	0
B_2Si	26%	1	0	0
B_1Fa	0	0	1	89%
B_2Fa	0	0	22%	1

Table 5 Tires: Demand Share Estimates (Averaged Across Stores)

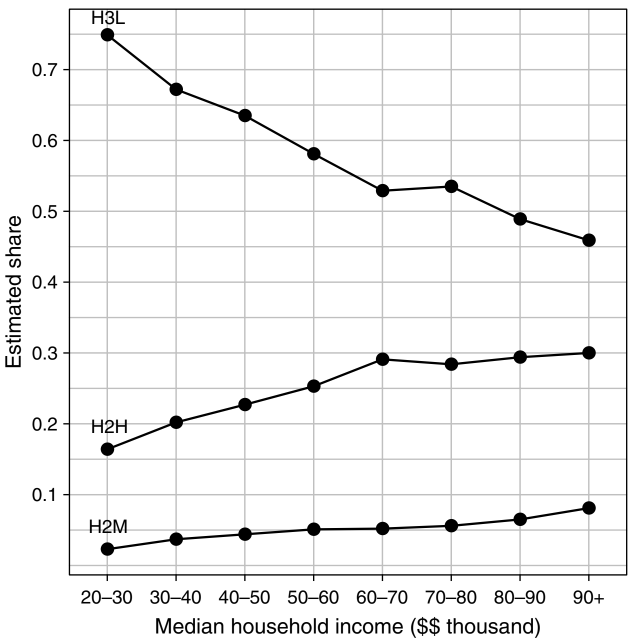
Brand-warranty	Sales share (%)	Demand share (%)
NH	1	4
NM	1	3
H1H	3	4
H2H	26	24
H2M	45	5
H3L	24	61

estimates were $\pi_S=2\%$, $\pi_L=6\%$, and $\pi_M=45\%$. Note in Table 3 that brand shares vary significantly between single serve and family size, showing the importance of our having combined brand and size into a single attribute, thus allowing for differences such as this. Similarly, as shown in Table 4, there is 89% substitution from brand 1 to brand 2 in family size, but only 18% in single serve.[4] The retailer found this the most interesting finding of the study, because they believed substitution rates varied, but had previously no way to measure the exact rates.

The most interesting finding for the tire retailer was that, as shown in Table 5, the demand share estimate for H3L is much higher than the sales share, whereas for H2M it is much lower. The reason for this appears to be that the retailer offered H3L in many fewer sizes than H2M; H3L is offered in only 15 of the 64 sizes, versus 52 sizes in which H2M is offered. But for those sizes where H3L and H2M are both offered, H3L outsells H2M by 40 to 1 on average, indicating that it is strongly preferred over H2M. The retailer offered H3L in fewer sizes because they preferred to sell the higher priced H2M and believed that their sales staff could convince customers to trade up. The substitution estimate of $\pi_M=45\%$ shows that many customers did trade up, and this explains the high sales share for H2M relative to its demand share. However, the 55% of the 61% of customers preferring H3L who did not substitute represents more than 34% of demand that was being lost due to the meager offering of H3L in the current assortment, suggesting that there was substantial opportunity to increase sales by reassorting.

[4] Beswick and Isotta (2010, p. 2) report a very similar finding for an orange juice study. For the leading brand, only 21% are willing to substitute to other brands, but for the second brand, 85% are willing to substitute.

Figure 2 **Tires: Share of H3L (H2H, H2M) is Negatively (Positively) Correlated with Income**



We can see that offering H3L in only a few sizes hurts revenue. The average prices of H3L and H2M in the sizes where both were offered were $28 and $36, respectively. Suppose that there were 100 consumers shopping the store and consider the two alternatives of offering H3L alone or H2M alone. Offering H2M would capture $(5\% * 100 + 45\% * 61\% * 100) * \$36 = \$1{,}168$ in revenues, whereas offering H3L would capture $(61\% * 100) * \$28 = \$1{,}708$, implying 46% additional revenue.

Figure 2 shows that the estimated share of H3L at each store is negatively correlated, and the share of H2H and H2M are positively correlated, with median income level in the zip code in which the store is located, which is intuitive and provides confirming demographic evidence to support the reasonableness of our parameter estimates.

In the tire application it was feasible for the retailer to offer store-specific assortments, since the tires are not actually displayed at the store. But the snack cake and appearance chemical retailers thought it would be administratively too complex to have more than, respectively, six and five different assortments.

For snack cakes, we applied our assortment heuristics for $L = 1, 2, \ldots, 6 \cup m$, keeping $K = 40$ for all stores.[5] Table 6 shows revenue as a function of L. We note that complete localization increases revenue to $8.11 million compared to the revenue of $7.38 million for $L = 1$.

[5] We only report results obtained using the forward greedy heuristic because the results based on the reverse greedy heuristic were not significantly different.

Table 6 **Snack Cakes: Impact of Amount of Localization on Revenue**

L	Revenues ($ million)
1	7.38
2	7.62
3	7.75
4	7.86
5	7.92
6	7.94
...	...
$m = 140$	8.11

However, 76.7% of this increase can be achieved with just six different assortments, suggesting that a small amount of localization can have a big impact.

For appearance chemicals, we worked first on a 2,183 store "warm-up" case and then the full 3,236 store case, generating up to five different assortments. Table 7 shows estimated revenue increases for various cases. Based on results of the 2,183 store case, the retailer concluded that at most three store clusters would be used, so these were the cases run for the 3,236 store case. The retailer liked the two-cluster solution and regarded cluster 1 as a base case that was representative of the chain as a whole and cluster 2 as a subset of stores differentiated, as shown in Table 8, by a higher level of tire related purchases, a preference for brand two, and a higher percentage of customers in a demographic the retailer called "urban/bilingual." This store segmentation was compelling for the retailer and agreed with more qualitative market research inputs they had received. Again we see that a small amount of localization produces most of the benefit.

To determine total market potential, we also ran an alternate scenario for appearance chemicals, expanding the potential SKU set to consist of all combinations of product attributes. This resulted in a chain optimal assortment lift of 15.2% and store optimal assortment lift of 19.7%, compared to the values of 11.8% and 14.2%, respectively, for the comparable cases using only SKUs available in the market.

5.6. Implementation

The tire and appearance chemical retailers implemented a portion of our recommendations, which allowed us to assess how our methodology performed under a stringent implementation test.

Table 7 **Appearance Chemicals: Estimated Revenue Increases (%) vs. Number of Store Clusters**

No. of clusters	2,183 stores (warm up)	3,236 stores (implementation)
1	14.1	11.8
2	14.4	11.9
3	14.6	12.0
5	14.9	

Table 8 Appearance Chemicals: Cluster Statistics for Top 25 SKUs (Based on Demand Estimates)

Revenue distribution by product type—package size, quality, and brand							
Segment—package size	Cluster 1 (%)	Cluster 2 (%)	Total (%)	Quality	Cluster 1 (%)	Cluster 2 (%)	Total (%)
Tire dressings/shines TRIGGER L	7.1	12.6	8.8	Good	12	14	13
Multipurpose protectants S	7.2	8.4	7.6	Better	60	67	62
Washes L	7.4	6.6	7.1	Best	28	20	25
Tire dressings/shines AEROSOL S	5.0	7.4	5.8	Total	100	100	100
Tire cleaners TRIGGER L	4.6	7.4	5.5				
Liquid Wax S	5.5	3.6	4.9				
Wash and Wax S	4.6	4.4	4.6	Brand 1	18	20	19
Washes S	4.9	3.8	4.6	Brand 2	7	11	8
Multipurpose protectants L	4.0	3.8	4.0	Brand 3	6	6	6
Tire foams (multipurpose) AEROSOL S	3.1	4.1	3.4	Brand 4	19	16	18
Carpet/upholstery cleaners AEROSOL S	3.2	3.0	3.1	Brand 5	12	7	10
Spray Wax S	3.0	2.6	2.9	Brand 6	6	7	7
Paste Wax S	3.0	2.2	2.8	Brand 7	2	2	2
Wheel care/cleaner. ALL TRIGGER S	2.7	2.8	2.7	Brand 8	11	10	11
Spray detailers L	3.0	2.1	2.7	Brand 9	19	21	19
Leather cleaners/conditioners TRIGGER S	2.7	2.0	2.5	Total	100	100	100
Wheel care/cleaner. ALL TRIGGER L	2.4	2.5	2.4				
Rubbing/polishing compounds S	2.1	1.5	1.9				
Tire dressings/shines BOTTLE GEL S	1.9	1.2	1.7		Demographic statistics		
Scratch removers S	1.8	1.1	1.6		Cluster 1	Cluster 2	Total
Rubbing/polishing compounds L	1.9	0.9	1.6	Income index	0.95	0.80	0.90
Glass cleaners TRIGGER L	1.5	1.6	1.5	% Suburban	85%	62%	78%
Plastic/lens cleaners, polishes and repair S	1.6	1.3	1.5	% Urban/bilingual	16%	42%	24%
Glass cleaners AEROSOL S	1.6	1.4	1.5	Store count	2,263	973	15%

Note. S, small; L, large.

Most of the products offered by the tire retailer were private label products. The retailer had relations with a number of suppliers from whom they could request production of a particular tire, but it was not always possible to find a supplier that would fulfill a given order. Usually the reason for declination was that the order was too small. The retailer asked us to optimize the assortment based on all attribute combinations so they could take the resultant "shopping list" to the market to secure as many as possible of our recommended new tires. Our optimized assortment deleted 47 SKUs and replaced them with 47 new SKUs. Of these 47 new SKUs, the retailer could find suppliers for 11, and these were added to the assortment in all stores, 10 H3L SKUs and 1 H1H.

Estimating the change in revenue from the current to the implemented assortment was complicated because, in addition to adding 11 new SKUs to the assortment, the retailer deleted more than 11 SKUs in each store. The number of SKUs deleted varied somewhat by store, but averaged 24 SKUs deleted. To achieve a fair comparison, in a store where N_s SKUs had been deleted and 11 added, we used the greedy heuristic to choose N_s minus 11 SKUs that were in the current assortment but not in the implemented assortment, and added them to create a modified implementation assortment that had the same SKU count as the preimplementation assortment at each store, which was used in evaluation. We then used parameters estimated for the calibration, validation, and implementation periods to estimate revenue for the current assortment, the modified implementation assortment, and the two optimized assortments. Table 9 gives revenue estimates and the percentage improvement over the current baseline assortment for these periods, and shows a 5.8% revenue increase for the new assortment.

Table 9 Tires: Revenue Estimates for Current, Implemented, and Optimized Assortments

	Parameter estimates used		
Revenues ($$ million) assortment	Jul–Dec 2004	Jan–Jun 2005	Jul–Dec 2005
Current assortment (Jul–Dec 2004)	80.2	74.9	72.3
Modified implementation (Jul–Dec 2005)	90.7 (13.1)	84.7 (13.1)	76.5 (5.8)
Recommended optimal assortment, chain	104.1 (29.8)	94.3 (25.9)	81.4 (12.6)
Recommended optimal assortment, store	108.2 (34.9)	99.2 (32.4)	83.6 (15.6)

Note. Percentage improvement over current revenues given in parentheses.

A 5.8% improvement is significant. For example, Canadian Tire reports achieving a 3%–4% annual revenue increase in existing stores during 2005–2009 and is targeting the same increase through 2012 as a result of all efforts, including assortment modification, to increase sales in existing stores.

The appearance chemical retailer also adopted our recommendations, using the two-cluster solution. In our recommended assortment, 20 SKUs in cluster 1 with the lowest estimated revenue were replaced by 20 new attribute combinations. In cluster 2, 19 existing SKUs were replaced. The retailer used exactly the assignment of stores to clusters in our recommendations. They also adopted our recommendations on which attribute combinations to add to the assortment, although in some instances more than one SKU in the market corresponded to the same attribute combination, with the result that the number of SKUs added exceeded the number of attribute combinations added. Twenty-two SKUs were added to cluster 1 and 25 were added to cluster 2. The choice of which SKUs to delete differed from our recommendations in the number of SKUs deleted and in which SKUs were deleted. Seventeen SKUs were deleted from cluster 1 and 23 were deleted from cluster 2. The choice of which SKUs to delete was guided by factors other than year to date revenue; for example, one car wash SKU was deleted due to a history of quality issues. In addition to assortment changes, their localization efforts included giving more prominent display and signage for tire products and brand 2 in the cluster 2 stores.

To evaluate the impact of these changes, we had available sales by cluster of the new assortment for the 27-week period January 1–July 8, 2010, and of the previous assortment for a comparable period in 2009. In the discussion below, we refer to these as 2010 and 2009 sales, respectively, while recognizing they were for only a portion of these years. SKUs can be segmented into three groups: kept SKUs that were in both the 2009 and 2010 assortments, deleted SKUs that were in the 2009 assortment but not the 2010 assortment, and added SKUs that were in the 2010 assortment but not the 2009 assortment. To evaluate the impact of the assortment changes, we compared 2010 kept plus added revenue to 2010 kept revenue plus an estimate of what 2010 revenue would have been for the 2009 SKUs deleted. Revenue is affected by a variety of factors other than assortment, including weather, the economy, and competitive activity. We measured the impact of these other factors by the ratio of the 2010 to 2009 revenue for kept SKUs and estimated the 2010 revenue of the deleted SKUs as their 2009 revenue times this factor. We are thus comparing the new assortment revenue to an estimate of what the old assortment would have sold in 2010.

We needed to deal with the fact that more SKUs were added than deleted. Twenty-two SKUs were added to cluster 1 versus 17 deleted, and 25 SKUs were added to cluster 2 versus 23 deleted. The retailer had a fixed amount of shelf space allocated to this category and accommodated the increase in SKU count by reducing the shelf space assigned to some of the existing SKUs. They therefore did not view the increase in SKU count as a cause for concern. Still, reducing the space for some existing SKUs might have caused greater stock-outs, reducing revenue in a way we could not capture. Thus, to make a more rigorous evaluation of benefits, we used the 22 and 25 SKUs for clusters 1 and 2, respectively, with lowest revenue in the 2009 evaluation period in estimating deletion revenue, thereby equalizing the add and delete counts.

The newly added SKUs were introduced some time after January 1, 2010, and hence were not on sale for the entire January 1–July 8, 2010, period and, moreover, took some time to build to a steady-state level of sales. Examining the weekly sales data of the added SKUs, we observed that it took them seven weeks to achieve a steady-state sales rate. Hence, we used added SKU revenue for weeks 8–27 scaled by 27/20 as our estimate of added SKU revenue for the period January 1–July 8, 2010.

The result of these calculations showed a 3.6% revenue increase due to the revised assortment. In addition, there may have been some improvement due to the localized product display and signage in cluster 2 stores that we were not able to measure. The retailer's appearance chemical team agreed with our assessment of benefits and believed that the reassortment exercise had been a success. We note that a 3.6% revenue increase is large relative to the usual impact of all things retailers do to increase same store sales. For example, in their annual reports, four leading auto parts retailers (Advance Auto Parts, Auto Zone, O'Reilly, and Pep Boys) reported comparable store sales increases over 2006–2010 that averaged 1.9%.

5.7. Forecasting New SKUs

One new SKU was added to the snack cake assortment in the July–December 2007 validation period. Eleven SKUs were added to the tire assortment and 25 to the appearance chemical assortment during the implementation periods. Table 1 shows the mean absolute percentage error (MAPE) of all new SKU forecasts. We note that these values compare favorably to the 30.7% MAPE for chain sales of two new SKUs reported by Fader and Hardie (1996).

Sources of forecast error include random fluctuation in sales, the approximation of representing SKU shares as the product of attribute shares, demand trends over time, and price changes. Table 10 shows how changes in relative prices across tire brand-warranties

Table 10 Tires: Price Changes and Impact on Demand Shares for a Representative Store

	Share of demand (price in $$)		
Brand-warranty	Jul–Dec 2004	Jan–Jun 2005	Jul–Dec 2005
NH	2 (73.9)	5 (69.7)	8 (77.6)
NM	2 (58.9)	3 (51.4)	6 (50.9)
H1H	3 (59.8)	10 (61.1)	8 (58.6)
H2H	16 (49.6)	21 (53.5)	22 (56.5)
H2M	7 (43.3)	10 (45.7)	29 (46.5)
H3L	70 (30.1)	51 (32.2)	27 (37.9)
H2M–H3L % price difference	43.7	41.9	22.9

relates to systematic changes in their demand shares over time. As the price difference between H2M and H3L narrowed from 43.7% to 41.9% to 22.9% over the three periods, the demand split between these two brand-warranties shifted from 7%/70% to 29%/27%.

6. Factors Influencing Localization Revenue Lift

The localization lift, defined as the revenue increase from using store specific assortments versus a single assortment for the chain was 12.2%, 5.8%, and 2.4% for the snack cake, tire, and appearance chemical applications. As we sought to understand what features of the problem data cause these differences in localization lift, the obvious reason was that some SKUs are more popular in some stores than in others. Although this is a necessary condition for a revenue increase from localization (clearly if all SKUs sold in the same proportions across all stores, there would be no benefit from localization), it turns out it is not a sufficient condition. For example, in the appearance chemical application, the single best selling SKU accounted for between 5.5% of revenue in its worst selling store and 16.6% of revenue in its best selling store, a 3.3 to 1 difference. Yet, even in the worst selling store, with a revenue share of 5.5%, it clearly made sense to have this SKU in a revenue maximizing assortment for this store, and hence this difference in sales rate had no impact on the localization lift. More generally, we can segment SKUs into three groups: (1) those with such high demand that they were carried in every store optimal assortment, (2) those with such low demand that they were in no store optimal assortment and (3) the remainder. Although there may be substantial variation in demand across stores for the first groups, none of this variation impacts localization lift.

We also noticed substantial variation in the breadth of assortment carried, from 40 of 92 possible SKUs for snack cakes to 130 of 154 possible SKUs for appearance chemicals. A broader assortment means that a single chain optimal assortment captures a greater fraction of potential demand, leaving less room for improvement from assortment localization.

We defined three metrics that we hypothesize could be correlated with localization lift. The coefficient of demand variation across all SKUs, COV-all, is defined as $\sqrt{\sum_{i\in N}\sigma_i^2}/\sum_{i\in N}\mu_i$, where μ_i and σ_i denote the mean and standard deviation, respectively, of revenue shares of SKU i across all stores. Motivated by the above observations, we also define COV-select (K) as the coefficient of variation for SKUs that are in some but not all store optimal assortments. This metric is a function of K because store optimal assortments depend on K. Chain optimal share (K) is the share of the total potential revenue achieved by a single, chain optimal assortment with K SKUs. So far we have not discussed the impact of substitution. We simply note that an increase in willingness to substitute increases chain optimal share (K) and thus decreases localization lift, which is intuitive, since if customers are more willing to substitute, there is less need to localize the assortment to their ideal tastes.

Table 11 gives localization lift and the three metrics for the three applications. We observe that the highest localization lift for cakes can be explained by the low value of chain optimal share (K) and high values of COV-all and COV-select (K). We also note that the difference in localization lift between tires and appearance chemicals can be explained by a high value of chain optimal share (K) and a relatively low value of COV-select (K) for appearance chemicals.

We also used the solutions to the three applications for K varying from 1 to n to create Figure 3, showing localization lift versus the percentage of maximum total SKUs in the assortment and chain optimal share (K). We note that localization lift varies with chain optimal share (K) as we have hypothesized. The dip in the lift index initially in Figure 3 is because the top SKUs at the chain level have different rank ordering across stores.

To further investigate the effect of demand variation and share captured by the chain optimal assortment on localization lift, we used the existing data to create 100 additional problem instances for each of the three applications by randomizing (a) the number of stores in the chain, (b) the actual stores sampled, and (c) K, the maximum number of SKUs allowed in the assortment. Table 12 gives the range used for each application in randomly generating store count and K. We then calculated localization lift, COV-all, COV-select, and chain optimal share (K) and regressed localization lift

Table 11 Explaining Localization Lift

Category	Lift	COV-all	COV-select (K)	Chain optimal share
Cakes	0.122	0.17	0.23	0.71
Tires	0.058	0.11	0.12	0.80
Appearance chemicals	0.024	0.10	0.10	0.93

Figure 3 Localization Lift vs. Maximum Total SKUs in the Assortment and Chain Optimal Share

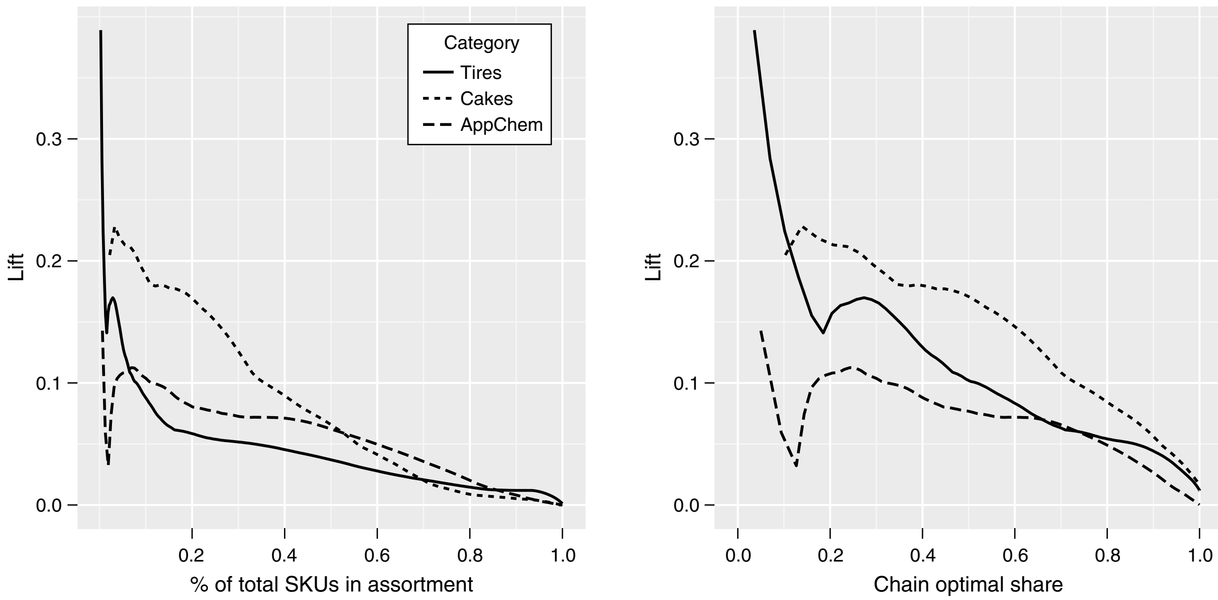


Table 12 Regression of Localization Lift

Coefficient	Cakes	Tires	Appearance chemicals
(Intercept)	0.147***	0.113***	0.186***
Chain optimal share	−0.095***	−0.166***	−0.157***
COV-select	0.045*	0.475***	0.339***
COV-all	−0.024	0.232	−0.000
Simulation details			
K (min–max)	10–40	10–100	5–40
Stores (min–max)	10–30	20–60	80–100
No. of instances simulated	100	100	100

against the three dependent variables; the results are summarized in Table 12. The regression results confirm our hypothesis that localization lift depends mainly on chain optimal share (K) and COV-select, which are highly significant, and not on COV-all, which is insignificant.

7. Conclusions

We have formulated a process for finding optimal assortments, comprised of a demand model, an estimation approach, and heuristics for choosing assortments. We apply this process to real data from three applications and show that it produces accurate forecasts for new SKUs and revenue increases of 5.8% and 3.6% in the two applications in which our results were implemented, increases that are significant relative to typical comparable store increases in these product segments.

We note the following observations from this research.

1. Forecast accuracy for new SKUs compared favorably to prior research in Fader and Hardie (1996) and was adequate to achieve significant benefits in implementation.

2. Sales are not true demand, but demand distorted by the assortment offered. We do not see demand for SKUs not offered, and the sales of SKUs offered may be increased above true demand due to substitution. The impact of these effects can be significant. In the tire application, the lowest price brand-warranty level had a demand share of 61%, but a sales share of only 5%, because the retailer offered the lowest price brand-warranty in few sizes. Adding more of this brand warranty to the assortment was a big source of the revenue increase attained.

3. Substitution can be measured, can vary significantly, and can have a major impact on the optimal assortment. In the snack cake example, in the family size, the probability of substituting from brand 1 to brand 2 was 89%, versus only a 22% probability of substituting from brand 2 to brand 1. This resulted in a complete replacement of brand 1 by brand 2 in the family size of the optimized assortment.

4. We use demographic data to confirm our parameter estimates in the tire and appearance chemical examples. The share of the lowest price tire and unwillingness to substitute up to a higher price tire were correlated with median income in the store area. In some instances, we also used demographic data to assist in estimating parameters.

5. The benefit from localizing assortments by store varied considerably, from 2% to 12%. We defined three metrics to understand the drivers of localization benefit: COV-all, the coefficient of variation across stores of all the SKUs; COV-select (K), the coefficient of variation for SKUs that are in some but not all store optimal

assortments; and chain optimal share (K), the share of the total potential revenue achieved by a single, chain optimal assortment with K SKUs. We used real data to randomly generate a large sample of problems, found the localization lift for each, and regressed against these three metrics. Only the second and third were statistically significant in explaining localization lift.

6. A limited amount of localization can capture most of the benefits of maximum localization. In the snack cake example, going from 1 assortment to 6 provided 77% of the benefit of going from 1 assortment to 140 assortments.

7. There may be interaction between attribute levels not captured by our demand model. In the case of tires, the demand for the least expensive brand-warranty level will be higher for a size tire that goes on an older, inexpensive car than for a tire that goes on a new, luxury car. We showed that this could be incorporated into our approach through latent class analysis.

Acknowledgments

The authors thank Abba Krieger and David Bell from the Wharton School for numerous valuable comments and helpful suggestions with this research, and Paul Beswick (partner and head, Oliver Wyman North American Retail Practice), Robert DiRomualdo (former CEO, Borders Group), Kevin Freeland (COO, Advance Auto), Matthew Hamory (principal, Oliver Wyman North American Retail Practice), Herbert Kleinberger (principal, ARC Consulting), Chris Morrison (senior VP of sales, Americas, Tradestone), Robert Price (chief marketing officer, CVS), and Cheryl Sullivan (vice president of product management, Revionics, Inc.) for helpful discussions on assortment planning practice. The authors also thank Carol Jensen, Tamara H. Jermyn, and Amanda O'Brien from Wawa, Inc. for support. The financial support of the Wharton School Fishman Davidson Center for Service and Operations Management and the Natural Sciences and Engineering Research Council of Canada [Grant 386497-12] are gratefully acknowledged.

References

Anupindi R, Dada M, Gupta S (1998) Estimation of consumer demand with stock-out based substitution: An application to vending machine products. *Marketing Sci.* 17(4):406–423.

Bell DR, Bonfrer A, Chintagunta PK (2005) Recovering SKU-level preferences and response sensitivities from market share models estimated on item aggregates. *J. Marketing Res.* 42(2):169–182.

Belloni A, Freund R, Selove M, Simester D (2008) Optimizing product line designs: Efficient methods and comparisons. *Management Sci.* 54(9):1544–1552.

Beswick P, Isotta M (2010) Making the right choices: SKU rationalization in retail. Working paper, Oliver Wyman, New York.

Chong JK, Ho TH, Tang CS (2001) A modeling framework for category assortment planning. *Manufacturing Service Oper. Management* 3(3):191–210.

Fader PS, Hardie BG (1996) Modeling consumer choice among skus. *J. Marketing Res.* 33(4):442–452.

Fisher M, Raman A (2010) *The New Science of Retailing: How Analytics are Transforming the Supply Chain and Improving Performance* (Harvard Business School Press, Boston).

Gaur V, Honhon D (2006) Assortment planning and inventory decisions under a locational choice model. *Management Sci.* 52(10):1528–1543.

Green PE, Krieger AM (1985) Models and heuristics for product line selection. *Marketing Sci.* 4(1):1–19.

Guadagni PM, Little JDC (1983) A logit model of brand choice calibrated on scanner data. *Marketing Sci.* 2(3):203–238.

Hotelling H (1929) Stability in competition. *Econom. J.* 39(153):41–57.

Kamakura WA, Russell GJ (1989) A probabilistic choice model for market segmentation and elasticity structure. *J. Marketing Res.* 26(4):379–390.

Kök G, Fisher ML (2007) Demand estimation and assortment optimization under substitution: Methodology and application. *Oper. Res.* 55(6):1001–1021.

Kök G, Fisher ML, Vaidyanathan R (2009) Assortment planning: Review of literature and industrial practice. Agrawal N, Smith SA, eds. *Retail Supply Chain Management* (Springer, New York), 99–153.

Lancaster K (1975) Socially optimal product differentiation. *Amer. Econom. Rev.* 65(4):567–585.

Mahajan S, van Ryzin G (2001) Stocking retail assortments under dynamic consumer substitution. *Oper. Res.* 49(3):334–351.

McGregor J (2008) At Best Buy, marketing goes micro. *BusinessWeek* (May 14), http://www.businessweek.com/stories/2008-05-14/at-best-buy-marketing-goes-micro.

O'Connell V (2008) Reversing field, Macy's goes local. *Wall Street Journal* (April 21), http://online.wsj.com/news/articles/SB120873643128029889.

Rosen S (1974) Hedonic prices and implicit markets: Product differentiation in pure competition. *J. Political Econom.* 82(1):34–55.

Smith SA, Agrawal N (2000) Management of multi-item retail inventory systems with demand substitution. *Oper. Res.* 48(1):50–64.

van Ryzin G, Mahajan S (1999) On the relationship between inventory costs and variety benefits in retail assortments. *Management Sci.* 45(11):1496–1509.

Vulcano G, Van Ryzin G, Ratliff R (2012) Estimating primary demand for substitutable products from sales transaction data. *Oper. Res.* 60(2):313–334.

Zimmerman A (2006) To boost sales, Wal-Mart drops one-size-fits-all approach. *Wall Sreet Journal* (September 7), http://online.wsj.com/news/articles/SB115758956826955863.

Zimmerman A (2008) Home Depot learns to go local. *Wall Street Journal* (October 7), http://online.wsj.com/news/articles/SB122333593880409529.