An Algorithm and Demand Estimation Procedure for Retail Assortment Optimization

Marshall Fisher

OPIM Department, The Wharton School, 3730 Walnut Street,

Philadelphia, PA 19104, USA

fisher@wharton.upenn.edu

Ramnath Vaidyanathan

Desautels Faculty of Management, 1001 Sherbrooke Street West,

Montreal, QC H3A 1G5, Canada

ramnath.vaidyanathan@mcgill.ca

June 2009, Revised May 2011

Abstract

We consider the problem of choosing, from a set of N potential SKUs in a retail category, K SKUs to be carried at each store so as to maximize revenue or profit. Assortments can vary by store, subject to a maximum number of different assortments. We introduce a model of substitution behavior, in case a customer's first choice is unavailable and consider the impact of substitution in choosing assortments for the retail chain. We view a SKU as a set of attribute levels, 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. 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 cakes, tires and automotive appearance chemicals applications, respectively with MAPEs of 16.2%, 19.1% and 28.7%.

1 Introduction

A retailer's assortment is the set of products they carry in each store at each point in time. Retailers periodically review and revise the assortment for each category of products they carry 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 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.

If a customer's most preferred product is not in a retailers assortment, they may elect to buy nothing or to purchase another product sufficiently similar to their most preferred product that they are willing to buy it. This possibility of substitution must be taken into account in both assortment optimization and in estimation. In estimation, 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 didn't find their preferred SKU in the assortment.

Most retailers use the same assortment for all stores, except that in smaller stores they might eliminate some SKUs. Recently however, localizing assortments by store or store cluster has become a high priority for many retailers. For example, Zimmerman (September 7, 2006), O'Connell (April 21, 2008), McGregor (May 15, 2008) and Zimmerman (October 7, 2008) describe recent efforts by Wal-Mart, Macy's, Best Buy and Home Depot to vary the assortment they carry at each store to account for local tastes. In the extreme, a retailer might carry a unique assortment in each store, but most retailers claim that this is administratively too complicated. For example, 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. Despite the flurry of interest in localization reported in the business press and which 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.

The assortment planning process varies greatly across different retailers and product categories. Retail product categories are commonly segmented into into apparel, grocery, and everything else, usually called hard goods.[1] An analytic approach to apparel assortments is challenging because rapidly changing tastes make sales history of limited value. Assortment planning is most developed in the grocery segment (where it is usually called category management), due in part to Nielsen/IRI who enlists households to record over time their grocery purchases in all stores. Much academic research on grocery consumer behavior has relied heavily on national panel data. Among other things, the data allows one to model substitution behavior by observing what a customer buys, if anything, when a product they purchase every week is unavailable due to a stockout. Many grocery retailers are now engaged in SKU rationalization efforts aimed at reducing SKU count with minimal impact on revenue.

The approach described in this paper best fits hard goods, where many retailers conduct an annual review of their various categories aimed at identifying SKUs to delete and add to the assortment in

[1]This discussion is based on Fisher and Raman (2010) and conversations with several retail executives including 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.

each category. Deletion decisions are easier, since sales data is available to indicate the popularity of existing SKUs, but current industry practice (for example, the household purchase data available in grocery is not available in hard goods) provides little if any hard data with which to forecast the sales of SKUs that might be added to the assortment, and hence a category manager is forced to rely on intuition and the representations of suppliers as to the merits of new products they are introducing. This also makes it impossible to forecast the revenue impact of strategic changes such as assortment localization.

It is apparent that the ability to forecast store-SKU demand for all potential SKUs, including those with which a retailer has no prior sales experience, and to intelligently localize assortments by store, would be valuable enhancements to current assortment planning practice, and the goal of this paper is to provide those enhancements.

Our approach follows the marketing literature in viewing a SKU as defined by 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 also estimate these substitution probabilities. We apply various heuristics to these demand and substitution estimates to determine optimized assortments. Our process can control the degree of localization by limiting the number of different assortments to be any level between a single assortment for the chain to 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.

We do not consider inventory decisions, so our approach fits when inventory decisions are unrelated to assortment decisions as is the case for many slow movers, where the retailer carries a small amount of inventory, often just a single unit, e.g jewelry, auto parts, books and CDs. The convenience store retailer in our study carried a single facing of each product and the tire retailer four of each tire.

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 model for assortment planning, describe a methodology for estimating parameters and optimizing assortments and test the process on real data. While these papers are an excellent start, there is obviously a need for much more research on the vast topic of assortment planning. This paper extends these papers, and hence the exisiting literature, in four ways.

1. We provide a thorough treatment of the important emerging topic of assortment localization. We allow a constraint on the number of different assortments so as to bound the administrative cost of localization and measure how the amount of localization impacts revenue. We compare

and explain differing levels of localization benefits across our three applications. Chong et al. (2001) don't deal with localization. Kök and Fisher (2007) allow a unique assortment for each store but don't 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) don't explicitly forecast new SKUs, although it would appear from their process that they could derive forecasts for new SKUs. However, they need consumer-level transaction data over multiple shopping trips and this detailed data is not available in most non-grocery applications; the standard data available is store-SKU sales data. 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 not previously considered in assortment decision support research. Both Chong et al. (2001) and Kök and Fisher (2007) use the multinomial logit demand model, which implicitly assumes that substitution demand is divided over available products in proportion to those products' market share. This assumption fits a situation in which products are similar to each other, but vary in a taste parameter, such as different flavors of yogurt or colors of apparel. By contrast, our demand model is a variant of the locational choice model which fits a situation in which some products are better substitutes for a given product than others. This was a very real feature of our applications; for example, the natural substitutes for a 14 inch tire are other 14 inch tires, not 15 inch tires. As another example, in our snack cakes application we found the probability of substituting from Brand 1 to Brand 2 was 89%, but only 22% of substituting from Brand 2 to Brand 1. Hence Brand 2 customers were much more loyal.

   The multinomial logit approach better fits some real assortment problems and the locational choice approach fits others better, so both approaches are needed. Our providing an approach based on a locational choice demand model is analogous to the way in which, for insight models, Gaur and Honhon (2006) provided a locational choice complement to the multinomial logit based analysis of van Ryzin and Mahajan (1999).

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

In Section 2, we review relevant prior literature. Section 3 provides our demand model and a formulation of the assortment optimization problem. Section 4 describes estimation of the demand model and heuristics for assortment selection. Section 5 presents results for the three applications. Section 6 analyzes the application results to understand differences in localization benefit and to assess the performance of the heuristics. Section 7 offers some concluding remarks.

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. SeeKök et al. (2008) for a more extensive review.

Consumer choice models constitute the fundamental platform for assortment planning, and may be classified as (1) utility based models, and (2) exogenous demand models. Utility based models assume that every customer associates a utility U_i with each product $i \in N$. 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.

The Multinomial Logit (MNL) model (Guadagni and Little 1983) assumes that the utilities U_i can be decomposed into a *deterministic* component u_i that represents the average utility derived by the population of customers, and a *random* component ξ_i that represents idiosyncrasies across customers. The ξ_i are assumed to be identical and independent Gumbel random variables with mean zero and scale parameter μ. Under these assumptions, the market share for each SKU $i \in A$ can be written as $\frac{\exp\left(\frac{u_i}{\mu}\right)}{\sum_{i \in A} \exp\left(\frac{u_i}{\mu}\right)}$.

In the Locational Choice model (Hotelling 1929, Lancaster 1966, 1975), every SKU $i \in N$ is represented as a bundles of attribute levels $z_i = (z_i^1, z_i^2, ..., z_i^T) \in \mathbb{R}^T$. Each consumer has an ideal point $y \in \mathbb{R}^T$ that defines his most preferred attribute levels. The utility this consumer associates to SKU i is $U_i^y = c - \tau \|y - z_i\|$, where c is the utility derived from his most preferred product y, and τ is the disutility associated with each unit of deviation from y. A consumer not finding his ideal product y in the assortment A substitutes the variant $j \in A$ that is located closest to his ideal point in the attribute space if $U_j^y > 0$, or declines to purchase if $U_j^y \leq 0$.

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 and they buy that product if it is in the assortment. They also all assume that if a customer's favorite product is not in the assortment, they may substitute a different product. Where the models differ is in their assumptions about substitution behavior. The exogenous demand model is the most flexible model, allowing for any substitution structure, but it has many parameters and is hence difficult to estimate in practice. The MNL model assumes that the demand for a missing product which 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. This assumption of the Locational Choice model most closely fits our applications. For example, in the tire assortment application, a 14" diameter tire is more like other 14" tires than it is like 15” tires. Accordingly, our demand model will be similar to the Locational Choice model, but with several adaptations to make it suitable for real applications.

Demand estimation has been studied by Fader and Hardie (1996). They use maximum likelihood 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. Estimation is parsimonious as they only need to estimate attribute-level utilities, and their model has the added advantage of being able to estimate demand of new products. Bell et al. (2005) describe a method for obtaining (SKU)-level preferences from estimated attribute level parameters, circumventing the need for direct estimation of the more complex SKU-level model. Anupindi et al. (1998) estimate demand and substitution probabilities for two products using sales transaction data from vending machines. Vulcano et al. (2009) use the *Expectation-Maximization (EM)* algorithm of Dempster et al. (1977) to develop a procedure to estimate demand from sales transaction data, when the underlying substitution is governed by a MNL model. Kök and Fisher (2007) and Chong et al. (2001) also estimate demand as part of an assortment optimization process, and these papers are discussed below.

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.

The stylized model research began with a pioneering paper by van Ryzin and Mahajan (1999). They study an assortment planning problem under a MNL consumer choice model and show that the optimal assortment consists of a certain number of the highest utility products. Mahajan and van Ryzin (2001) study the same problem allowing for stock-out substitution and develop heuristics based on a sample path approach. Cachon et al. (2005), Caro and Gallien (2007), and Maddah and Bish (2007) extend the van Ryzin Mahajan model in various ways.

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, indicating maximum differentiation, with no substitution between products in the assortment. 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 a product line design problem in which there are m consumer segments indexed by i, and n products, indexed by j. Every consumer segment i has a utility associated with product j denoted by u_{ij}. A consumer chooses from all available products the one that maximizes his utility. Green and Krieger then formulate the problem of which k products out of the n should a firm select so as to maximize (1) consumer welfare or (2) firm profits, and propose solution heuristics. Green and Krieger (1987a,b, 1992), McBride and Zufryden (1988), Dobson and Kalish (1988) and Kohli and Sukumar (1990) extend this line of research. 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 based on the Guadagni and Little (1983) brand-share model. They use 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. Although their model can implicitly predict SKU level demand, their explicit focus is on brand-shares, and hence they don't create or measure the accuracy of SKU level demand forecasts. They report an average 50% mean squared error of predicting brand choice at the customer level across different product categories.

Kök and Fisher (2007) use an exogenous demand model to study a joint assortment selection and inventory planning problem in the presence of shelf-space constraints. They note that the constrained shelf space implies that average inventory per SKU is inversely related to the breadth of the assortment and assess the tradeoff between assortment breadth vs in-stock level for the SKUs carried. They consider substitution and assume that substitution demand accrues to available products in proportion to their original market share, as in the MNL model. They provide a process for estimating demand and substitution rates and apply their method to data from 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. For concreteness, we assume that the goal is to maximize revenue, since this was the primary concern in our three applications, although it is straightforward to adapt our approach to maximizing other functions, such as unit sales or dollar gross margin. We define the following parameters.

$N = \{1, 2, ..., n\}$	Index set of all possible SKUs a retailer could carry in this category
$M = \{1, 2, ..., m\}$	Index set of all stores
K	Maximum number of SKUs per assortment
L	Maximum number of different assortments
D_i^s	Number of customers at store s for whom SKU i is their most preferred product
p_i	Price of SKU $i \in N$

The parameters K and L would be specified by the retailer. For expositional simplicity, K does not vary by store, but it would be easy to modify our process to enable K varying by store, and in fact we do this in our computational work. L would be chosen to lie between 1 and m to tradeoff 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 $a(s)$ denote the index of the assortment assigned to store s. A solution to the assortment optimization problem is completely defined by the portfolio of assortments S_l, $l = 1, 2, ..., L$ and $a(s) \in \{1, 2, \ldots, L\}$, for all $s \in M$.

Our demand model assumes that a consumer shopping this category in store s has a most preferred SKU $i \in N$, but might be willing to substitute to other SKUs if $i \notin S_{a(s)}$. 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.

A	Number of attributes
a	Attribute type index
N_a	Number of levels of attribute a,$a = 1, 2, \ldots, A$
f^s_{au}	Fraction of customers at store s who prefer level u of the attribute a, $u = 1, 2, \ldots, N_a$, $a = 1, 2, \ldots$,A
π^s_{auv}	Probability that a customer at store s who's 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. For example, size is an attribute of a tire and a 14 inch tire is not a feasible substitute for a customer with a 15 inch wheel.

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 $\pi^s_{ij} = \Pi^{a=A}_{a=1} \pi^s_{ai_aj_a}$. 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 $j(i, S) = \arg\max_{i\in S} \Pi^A_{a=1} \pi^s_{ai_aj_a}$.

In using store sales data to estimate the parameters of our model,we first estimate D^s, the total unit demand in store s, as total unit sales divided by the share of demand captured, as defined by demand shares and substitution probabilities, and then set $D^s_i = f^s_i D^s$.

The revenue earned by store s using assortment $S_{a(s)}$ can then be written as

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

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 $S_l, \mid S_l \mid \leq K, l = 1, 2, \ldots, L$ and $a(s)$ for all $s \in M$ to maximize $\sum_{s\in M} R_s(S_{a(s)})$.

Our demand model most closely resembles the locational choice model, but with three important differences: (1) in the locational choice model, the probability of purchasing the closest substitute from the assortment is either 0 or 1 (depending on the no-purchase utility), while we allow general substitution probabilities, (2) consumers are assumed to be distributed in a continuous space in the locational choice model, while we allow customer locations to be restricted to discrete locations in the attribute space, and (3) all attribute levels in the locational choice model have a numeric value which allows the calculation of distance between products and identification of the nearest product to a given ideal point, whereas the nearest product to an ideal point in our model is identified via the substitution probabilities, which can be thought of as inducing a distance metric for attributes that can't be located in a space. These three enhancements were needed to make the locational choice model operational in the applications we consider.

We allow substitution probabilities to vary by store because we found in our applications that they did in fact vary by store. For example, we will see in Section 5 that the willingness of a consumer to substitute to a higher priced product varies by store and is 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 three-fold: (1) all prior publications we are aware of that use attributes in demand estimation make a similar assumption. Fader and Hardie (1996) is typical of the approach followed in the literature. They 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 "*In both the marketing and economics literature, it is common to assume an additive utility function*" 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 cakes application (Section 5.1), 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 (Section 5.2), the attributes are size, brand (4 brands) and mileage warranty (low, medium, high). 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 tires example, 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 six brand-warranty levels correspond to different price points and quality. There is a clear interaction between brand-warranty shares across price points and 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 homogenous segments (which are latent) and allowing the brand-warranty shares to be conditional on segment membership.

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 Section 5, we will exploit special structure of the applications to refine this approach. Let S denote the assortment carried in a particular store and x_i 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 $LH(f, \pi) = C \prod_{j \in S} \left[\frac{F_j(S)}{F(S)}\right]^{x_j}$ where the proportionality constant is $C = \dfrac{(\sum x_{j \in S})!}{\prod_{j \in S} x_j!}$. The maximum likelihood estimates (MLE) for the parameters (f, π) can be obtained by maximizing the log-likelihood function

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

subject to the constraints

$$\begin{aligned}
\sum_{u=1}^{N_a} f_{au} &= 1, \quad a = 1, 2, \ldots, A && (2)\\
f_i &= \Pi_{a=1}^{a=A} f_{ai_a} \quad i = 1, 2, \ldots, n && (3)\\
\pi_{ij} &= \Pi_{a=1}^{a=A} \pi_{ai_a j_a} \quad i = 1, 2, \ldots, n \text{ and } j = 1, 2, \ldots, n && (4)\\
f_{au}, \pi_{auv} &\in [0, 1] \quad \forall a, u, \text{ and } v && (5)
\end{aligned}$$

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} as

$$\begin{aligned}
f_{au} &= \frac{\exp(\widehat{f}_{au})}{\sum_{u=1}^{N_a} \exp(\widehat{f}_{au})}, \quad u = 1, 2, \ldots, N_a - 1 \text{ and } \widehat{f}_{aN_a} = 1 && (6)\\
\pi_{auv} &= \frac{\exp(\widehat{\pi}_{auv})}{1 + \exp(\widehat{\pi}_{auv})}, \quad \forall a, u, \text{ and } v && (7)
\end{aligned}$$

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 $D = \frac{\sum_{i \in S} x_i}{F(S)}$, and D_i as $f_i D$.

4.2 Estimating Prices for New SKUs

We endeavored to set prices on SKUs not currently carried by the retailer in a way that would 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:

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

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

This is a hedonic pricing equation, and has been extensively used in economics (Rosen 1974, Goodman 1998, Pakes 2003).

4.3 Heuristics for Choosing Assortments

If there were no substitution, then the assortment problem could be optimally solved by a greedy algorithm that chose SKU's in decreasing order of their revenue contribution. But substitution makes the objective function nonlinear, because the contribution of a SKU depends in part on it's substitution demand, which depends on which other SKUs are 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 $a(s), s \in M$ 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. Belloni et al. (2008) find that for the product line design problem, greedy and interchange together find 98.5% of optimal profits on average for randomly generated problems, and 99.9% for real problems.

We first define a greedy heuristic for finding a single assortment $S^G(T)$ for a specified subset of stores T. In the statement of Greedy(T) below, we define $R_s(\emptyset) = 0$.

Greedy(T)

1. INITIALIZE $S^0 = \emptyset$, $k = 1$
2. WHILE $(k \leq K)$ DO
   (a) $j_k = \arg\max_{j \notin S^{k-1}} \sum_{s \in T} R_s \left(S^{k-1} \cup \{j\}\right)$
   (b) $S^k = S^{k-1} \cup \{j_k\}$
   (c) $k = k + 1$
   (d) END WHILE
3. RETURN $S^G(T) = S^K$

We also use an interchange heuristic which starts with a given assortment and tests whether interchanging a SKU which 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 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 Greedy(T) $m+1$ times with $T = M$ and $T = \{s\}, s \in M$, initialize $S_1 = S^G(M)$, and assign all stores to this assortment. If $L > 1$, we identify the assortment $S^* \in E = \left\{S^G(s) \mid s \in M\right\}$ to add that would maximize the incremental revenue gain. To identify S^*, we calculate the incremental revenue gain from adding each assortment $S \in E$ by reassigning stores to their revenue maximizing assortment in $S_1 \cup S$ and calculating the increase in revenue due to the reassignment. We then choose as S^* the assortment that gives the greatest revenue increase in this process. 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 add to this portfolio the assortment that leads to the highest increase in revenue and reassign stores to the enhanced set of assortments.

A detailed specification of forward greedy is given below.

Forward Greedy for Finding L Assortments

1. INITIALIZE
   (a) $l = 1$, $S_1 = S^G(M)$
   (b) $\mathbf{C}^l = \{S_i \mid i = 1, 2, ..., l\}$
   (c) $\mathbf{E} = \left\{S^G(\{s\}) \mid s \in M\right\}$
   (d) $a(s) = 1$, $\forall s \in M$

2. WHILE ($l < L$) DO

   (a) $S^* = \arg\max_{S \in \mathbf{E}} \sum_{s \in M} R_s\left(S_{a^*(s)}\right) - R_s\left(S_{a(s)}\right)$
   where $a^*(s) = \arg\max_{i:S_i \in \mathbf{C}^l \cup S} R_s(S_i)$

   (b) $l = l + 1$

   (c) $S_l = S^G(\{s^*\})$

   (d) $a(s) = \arg\max_{i \leq l} R_s(S_i)$

   (e) $T_i = \{s \in M \mid a(s) = i\}, i = 1, 2, ..., l$

   (f) $S_i = A^G(T_i), i = 1, 2, ..., l$

   END WHILE

3. RETURN $A_l, l = 1, 2, ..., L$, $a(s), s \in M$

In the reverse greedy heuristic, we first apply Greedy(T) m times with $T = \{s\}$ for all $s \in M$, initialize $S_i = S^G(M)$ for $i = 1, 2, ..., m$ and set $a(s) = s$ for all $s \in M$. If $L < m$, we identify the single assortment $S^* \in \mathbf{E} = \{S_i \mid i = 1, 2, ..., m\}$ to delete that would minimize the revenue loss. We calculate the incremental revenue loss from deleting any assortment $S \in \mathbf{E}$ by reassigning stores to their revenue maximizing assortment in $\mathbf{E} - S$ 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. We omit a detailed statement of the reverse greedy due to space limitations.

5 Results

5.1 Regional Convenience Chain

This retailer 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 from a consumer's perspective was whether the size was single-serve or family size, and hence we grouped sizes into *Single Serve* (S) and *Family Size* (F). Further, 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 SB_1, SB_2, FB_1 and FB_2 in order.

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 purchasing from this category. Hence, we assumed that the probability of substituting across flavors is negligible and 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 $i(j)$ to be the brand-size that would substitute to j if $i(j)$ is not in the assortment and set $i(1) = 2, i(2) = 1, i(3) = 4$ and $i(4) = 3$. We need to estimate the 23 flavor shares f_{1v}, 4 brand-size shares f_{2b}, and 4 substitution probability parameters $\pi_{12}, \pi_{21}, \pi_{34}$ and π_{43}.

We used Store-SKU sales data for the six month period from July 2005 to December 2005 to estimate model parameters at each of the 140 stores. We describe here a refinement of the estimation procedure described in Section 4.1 that exploits some special structure of this application, namely that there are two attributes with no substitution across one of them. Estimation was done at the store level, but for simplicity in the discussion below, we drop the store subscript. As before, let x_j denote the total sales of SKU j at a particular store during July 2005 to December 2005 and let v_j and b_j denote the flavor and brand-size, respectively, of SKU j. Given an assortment S, the sales share of SKU $j \in S$ at any store can then be expressed as $F_j(S) = f_{1v_j} f_{2b_j}$, if $i(b_j) \in S$ and $F_j(S) = f_{1v_j} \left(f_{2b_j} + f_{2i(b_j)} \pi_{i(b_j)b_j} \right)$, if $i(b_j) \notin S$. We can then write Equations (1) - (5) from Section 4.1 as maximizing

$$LL(f, \pi) = \sum_{j \in S} x_j \log F_j(S) - \left(\sum_{j \in S} x_j \right) \log F(S) \tag{9}$$

subject to the constraints

$$\begin{aligned} \sum_{v=1}^{23} f_{1v} &= 1, \\ \sum_{b=1}^{4} f_{2b} &= 1 \end{aligned} \tag{10}$$

where all variables are $\in [0, 1]$.

The Lagrangian can be written as

$$\begin{aligned} H = & \sum_{j \in S} x_j \log F_j(S) - \sum_{j \in S} x_j \log F(S) - \lambda_1 \left(\sum_{v=1}^{23} f_{1v} - 1 \right) - \lambda_2 \left(\sum_{b=1}^{4} f_{2b} - 1 \right) \\ & - \sum_{b=1}^{4} \mu_b \pi_{i(b)b} - \sum_{b=1}^{4} \gamma_b (1 - \pi_{i(b)b}) \end{aligned} \tag{11}$$

Applying first-order conditions with respect to the flavor shares f_{1v}, we get

$$\frac{\partial H}{\partial f_{1v}} = \frac{1}{f_{1v}} \left[\sum_{j \in S, v_j = v} x_j - \frac{\sum_{j \in S} x_j}{F(S)} F_j(S) \right] - \lambda_1 = 0 \tag{12}$$

Multiplying Equation 12 by f_{1v}, and adding the equations across all values of v, gives us

$$\lambda_1 = \sum x_j - \frac{\sum_{j\in S} x_j}{F(S)} \sum_{j\in S} F_j(S) = 0 \tag{13}$$

Hence, we get

$$\frac{\sum_{j\in S, v_j = v} F_j(S)}{F(S)} = \frac{\sum_{j\in S, v_j=v} x_j}{\sum_{j\in S} x_j} \tag{14}$$

which on simplification yields

$$f_{1v} = \frac{\sum_{j\in S, v_j=v} x_j}{\sum_{j\in S} x_j} \frac{F(S)}{\sum_{j\in S, v_j=v} \frac{F_j(S)}{f_{1v_j}}} \tag{15}$$

Note that $\frac{F_j(S)}{f_{1v_j}} = f_{2b_j} + f_{2i(b_j)}\pi_{i(b_j)b_j}$, and can be expressed in terms of the brand-size shares and substitution probabilities, and $F(S)$ can be calculated by using the fact that $\sum_{v=1}^{23} f_{1v} = 1$. This reduces the number of parameters to be estimated from 31 to 8, as the log-likelihood function can now be written as a function of the brand-size shares (f_{2b}) and substitution probabilities ($\pi_{i(b)b}$) alone.

As was described in Section 4, we transform the variables using Equations (6) and (7) to impose the constraints that they lie between 0 and 1, and use numerical methods based on gradients from randomly generated starting points to maximize this log-likelihood function. Tables 1 and 2 show the average parameter estimates across all stores.

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 $\frac{\sum_{s\in M}\sum_{j\in S^s}\left|\frac{x_j^s}{\sum_j x_j^s} - F_j^s(S^s)\right| x_j^s}{\sum_{s\in M}\sum_{j\in S^s}\frac{x_j^s}{\sum_j x_j^s}}$. 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. The MAD for this retailer was calculated to be 16.4% at the Store-SKU level and 6.2% at the Chain-SKU level.

We used a hedonic regression as described in Section 4.2 to assign price to SKUs not currently offered. The regression R^2 is 85.5% and confirms our assumption that prices on existing SKUs were set to be correlated with attribute levels that determine consumer value. We multiplied the estimated prices by a scale factor of 0.97 so as to equalize the estimated revenue at the chain level to the actual revenue, which will facilitate comparison of new optimized assortments with current revenue.

To validate our results, we used Store-SKU sales data for the six month period from July 2007 to December 2007. For this period, we had data for only 54 of the 140 stores in the chain. We used the previously estimated parameters to compute the share of sales for each SKU for the validation period

(July 2007 to December 2007), at the store and chain levels and compared it with the actual sales shares. The sales-weighted MAD of predicted sales shares from the actual sales shares was calculated to be 40.1%[2] at the store-SKU level and 25.8% at the chain-SKU level.

One new SKU was added to the assortment in the July-December 2007 period, Butterscotch in Brand 2 Single Serve. The MAD and MAPE[3] of the predicted sales shares from the actual sales shares at the chain level for the newly introduced SKU was 16.2%. The 16.2% MAPE compares favorably to the 30.7% MAPE for chain sales of two new SKUs reported by Fader and Hardie (1996), the only prior reporting of which we are aware of the errors of forecasts for new SKUs based on sales data.

Sources of error affecting both the calibration and validation samples include random fluctuation in sales and the approximation of representing SKU shares as the product of attribute shares. Additional sources of error for the validation sample include changes in relative prices across brand-sizes, which affect share and substitution probability estimates, and a steady increase in the demand shares of some newer flavors.

Figure 1 shows the results of applying the forward greedy heuristic described in Section 4.3 to compute optimized assortments at the chain level, varying the number of SKUs in the assortment from 1 to 92. Figure 1 also shows results for the tires and appearance chemicals examples which are discussed in Sections 5.2 and 5.3. Figure 1 shows the percentage captured of the maximum possible revenue if all SKUs were offered, as a function of K expressed 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.

To quantify the potential improvement in revenue, we compare the assortments we generated for $L = 1$ and $L = m$ to the current assortment, which had a revenue of \$6.19 million. Because SKU count varied somewhat by store, in computing the revenue of the $L = 1$ assortment, we first generated the greedy solution for a SKU count equal to the maximum SKU count across all stores, and then for store s with SKU count K_s, we used the first K_s SKUs chosen by greedy. The revenue of this solution was \$8.01 million, a 29.2% increase over the current revenue. The revenue for store-specific assortments with SKU counts of K_s for each store was \$8.75 million, a 41.4% increase over the current revenue. We thus see that maximum localization adds 12.2% beyond the 29.2% achieved by chain level optimization.

These are estimates of revenue improvement based on the calibration sample. If these assortment changes were implemented, we would expect the actual improvement to be less because of differences in the calibration sample sales and sales during the period of implementation. To determine how much these estimated revenue improvements would be eroded during implementation due to forecast errors, we used parameter estimates based on the validation sample Store-SKU sales data (July 2007

[2]The 54 stores in this analysis had a higher MAD of 24.6% at the Store-SKU level and 10.3% at the Chain-SKU level in the calibration sample, as compared to 16.4% at the Store-SKU level and 6.2% at the Chain-SKU level for the whole chain. This suggests that if we use data for the whole chain, then the Store-SKU level MAD in the validation sample, may proportionally come down from 40.1% to 26.7%.

[3]We define MAPE as $\dfrac{\sum_{s \in M} \sum_{j \in S^s} \left| \dfrac{\frac{x_j^s}{\sum_j x_j^s} - F_j^s(S^s)}{\frac{x_j^s}{\sum_j x_j^s}} \right| x_j^s}{\sum_{s \in M} \sum_{j \in S^s} x_j^s}$

to December 2007) to estimate the revenue that would have been achieved had our recommended assortment changes been implemented in the validation period. Because we were only able to obtain data on 54 stores during our validation period (vs. 140 stores in the calibration period), we compared the revenue lift of these 54 stores based on the calibration sample with the lift based on the validation sample. The calibration period revenue of the current 54 store assortment was \$2.43 million and the store-optimal assortment estimated revenue based on the calibration period was \$3.44 million, which is a 41.5% increase. Recomputing revenue estimates for these two assortments during the validation period gives \$2.5 million and \$3.0 million, a 20% increase. We see that the revenue improvement has eroded by half because the calibration period parameter estimates are an imperfect representation of the validation period reality, although a 20% revenue increase is still economically very significant. The calibration period revenue of the single chain-wide assortment was \$2.74 million indicating that localization would lead to an incremental 12.8% gain in revenues. The corresponding number calculated based on the validation period was \$2.69 million indicating a localization lift of 7.6%.

Table 5 provides data for an optimized assortment at a representative store and shows that the assortment has an intuitive property; the higher the sales rank of a flavor, the more brand-sizes are carried. Moreover, the same brand-sizes were carried for all flavors that had the same number of brand-sizes. If one brand-size were offered, it is B1S, if two are offered they are B1S and B2F and if three offered, they are B1S, B2F and B2S.

The result that was initially most surprising to the retailer is that the optimal assortment completely drops *Brand 1* in the *Family Size.* This is easily explained by looking at Table 2 which shows that 89% of consumers are willing to switch from B_1 to B_2 in the family-size segment[4]. Hence, by not offering B_1F, which accounts for 6% of primary demand, the retailer only loses $6\% * 11\% \sim 0.7\%$ of demand, which is more than made up by carrying more brand-sizes in other flavors. This result made sense to the retailer, who told us that Brand 1 was strongest in single serve, but Brand 2 was by far the strongest in family size, and that's why so many customers were willing to substitute from Brand 1 to Brand 2. They found this the most interesting finding of the study, as they believed substitution rates varied, but had previously had no way to measure the exact rates.

While using a unique assortment from each store adds 12.2% to revenue, this retailer believed that it would be unmanageably complex to have more than 6 assortments for the chain, because for each assortment they needed to develop a diagram (called a Planogram) showing how the product would be displayed in the store.

To quantify the benefit of a realistic level of localization, we applied our assortment heuristics for $L = \{1, 2, ..., 6\} \cup \{m\}$, keeping $K = 40$ for all stores.[5] Table 3 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 $L = 1$. However, 76.7% of this increase can be achieved with just 6 different assortments, suggesting that a small amount of localization can have a big impact.

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

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

5.2 National Tire Retailer

Tire attributes include brand, size, mileage warranty, price, speed rating and load limit. However, these attributes are not independent of each other. For example, size is positively correlated with load limit, while mileage warranty is correlated with speed rating. Based on discussions with management and analysis of attribute data, we concluded that brand, size and mileage warranty were the fundamental defining attributes of a tire relevant to assortment planning.

The retailer offered several nationally advertised brands that they believed were equivalent to the consumer, and which we denote 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 (for example, national brands were always offered only with high or medium warranty, while $H3$ tires were offered only with low warranty) to identify the following six brand-warranty combinations: NH, NM, $H1H$, $H2H$, $H2M$, $H3L$. Sixty four distinct tire sizes were offered, resulting in $64 * 6 = 384$ distinct 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; for example, a 14" diameter tire cannot be used on a 15" wheel. Table 4 is based on estimates provided by the Vice President of the tire category for the retailer 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.*

We used sales data at the Store-SKU level for the six month period from July 2004 to December 2004 to fit the model and estimate its parameters. We need to estimate 64 size shares, 6 brand-warranty shares, and 3 substitution probabilities. The estimation procedure followed was equivalent to the procedure described for snack cakes if we view *Size* as being equivalent to *Flavor* (in that there is no substitution across *Size* or *Flavor*) and *Brand-Warranty* being equivalent to *Brand-Size.* In particular, we can express log-likelihood as a function of the brand-warranty shares and substitution probabilities and hence need to estimate only 6 brand-warranty shares and 3 substitution probabilities for each store.

This process worked for 319 of the retailer's 574 stores, but at 255 stores there was insufficient data to determine all 6 brand-warranty shares. In particular, in these stores there was no size in which brand-warranties $H2M$ and $H3L$ were both offered, so it was not possible to identify the split of demand between $H2M$ and $H3L$. From the parameter estimates in the 319 stores with sufficient data to estimate all parameters, we observed that the share of $H3L$ at a store is correlated with median household income ($R^2 = 0.15, p < 0.10$). We regressed the $H3L$ share against median income for the stores at which we could estimate all the parameters and used the regression estimate for the share of $H3L$ at other stores. We then used MLE to estimate the remaining parameters. Of the 255 stores with insufficient data, there were 52 stores where we could neither identify the share of $H3L$ nor $H2M$. For these stores, in addition to estimating $H3L$ shares, we also estimated $H2M$ shares by regressing it against income. As before, we used MLE to estimate the remaining parameters.

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 to be expected. In addition to supporting the parameter estimation process as described above, these results provide confirming demographic evidence to support the reasonableness of our parameter estimates.

Tables 6 and 7 show the average across all stores of the estimated brand warranty demand shares and substitution probabilities. The sales-weighted MAD of sales shares predicted based on the parameter estimates from the actual sales shares is 13.6% at the store-SKU level and 4.5% at the chain-SKU level.[6]

Table 6 also gives sales share estimates which can be compared with the demand share estimates. The most interesting comparisons are the demand share estimate for $H3L$, which is much higher than the sales share, and for $H2M$, which is much lower than the sales share. 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 $40 : 1$ 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 to this tire. The substitution estimate of 45% shows that many customers did in fact 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 re-assorting.

We can see that offering H3L in only a few sizes hurts revenue. The average price of $H3L$ and $H2M$ in the sizes where both were offered was \$28 and \$36 respectively. Suppose that there were 100 consumers

[6] As discussed in Section 3, there is some interaction between size and brand-warranty of a tire. One way to account for this interaction is to use a latent class model (Fader and Hardie 1996, Kamakura and Russell 1989). In a latent class model, we assume that there are several homogenous segments of sizes, and the brand-warranty shares vary across size segments, but are the same within each segment. For example, a size segment could include tires that fit old cars, and the brand-warranty shares reflect this in that the shares of less expensive tires are relatively higher. The estimation problem then reduces to maximizing the likelihood function by jointly estimating the probabilities of segment membership for each size and the associated brand-warranty shares, which can be achieved by using the Expectation Maximization algorithm. Each size is assigned to the size segment for which it has the highest segment membership probability. The choice of the optimal number of segments is made by the Bayesian Information Criterion (BIC), which penalizes the likelihood function for the addition of segments. We estimated a latent class model for a subset of 20 stores. We decided on using two size segments based on BIC. The average share of H3L was 34.2% and 71.3% for the two segments. We also find that the sales-weighted MAD of sales share estimates improved from 13.6% to 11.3%.

shopping the store and consider the two alternatives of offering $H3L$ alone or $H2M$ alone. Offering $H2M$ would capture $(5\% * 100 + 45\% * 61\% * 100) * \$36 = \$1168$ in revenues while offering $H3L$ would capture $(61\% * 100) * \$28 = \$1708$ implying 46% additional revenue.

We used the hedonic regression described in Section 4.2 to assign price to SKUs not currently offered. The regression R^2 is 96.32% which supports our assumption that prices on existing SKUs are based on attribute levels. As before, we multiplied the estimated prices by a scale factor of 1.05 so as to equate the estimated revenue of the current assortment to the actual revenue of this assortment.

To validate our results, we used Store-SKU sales data for the next six month period from January 2005 to June 2005. We used the previously estimated parameters to compute Store-SKU sales shares and compared them with actual sales shares for the validation period (January 2005 to June 2005). The sales-weighted MAD of predicted sales shares from the actual sales shares was 38.2% at the Store-SKU level and 21.1% at the chain-SKU level.

Figure 1 shows the results of applying the greedy heuristic described in Section 3 to compute optimized assortments at the chain level, varying the maximum number of SKUs in the assortment, and Table 8 summarizes the store-optimal assortment for a representative store in terms of additions and deletions to the current assortment.

To quantify the potential improvement in revenue, we compare the assortments we generated for $L = 1$ and $L = m$ to the current assortment, which had a revenue of \$80.2 million. Because SKU count varied somewhat by store, in computing the revenue of the $L = 1$ assortment, as in the snack cakes example, we first generated the greedy solution for a SKU count equal to the maximum SKU count across all stores, and then for store s with SKU count K_s, we used the first K_s SKUs chosen by greedy. The revenue of this solution was \$104.1 million, a 30.1% increase over the current revenue. The revenue for store specific assortments with SKU counts of K_s for each store was \$108.7 million, a 35.9%[7] increase over the current revenue. We thus see that maximum localization adds 5.8% beyond the 30.1% achieved by chain level optimization.

In contrast to the snack cakes example, where a unique assortment per store was not feasible due to planogramming complexities, it is completely feasible here for the retailer to offer store-specific assortments, since the tires are not actually displayed at the store. Hence, we did not compute revenue lifts for values of L between 1 and m as we did with snack cakes.

We also performed a sensitivity analysis to determine how robust our results were to the assumptions made about substitution probabilities. We varied the estimated substitution probabilities by increasing/decreasing them by a factor of 2 and computed the revenues obtained from the chain-wide optimal assortment. Our analysis revealed that the optimal assortment and sales lift were sensitive to only one substitution parameter, the substitution probability from $H3L$ to $H2M$. This makes sense because the demand shares of other brand-warranties are low, and the fraction of customers who substitute if these low share options are not offered has little impact on revenue. Table 9 shows how the lift

[7] We also optimized the assortment to maximize gross margins using approximate gross margin data by brand. The increase in gross margins was 32%, still a significant number.

in revenues varied as a function of this average substitution probability. Note that for the base case, where the average substitution probability from $H3L$ to $H2M$ is 0.45, we obtain a 30.1% increase in revenues, whereas when this probability is 1, then the increase reduces to 9%. We can interpret this 9% as the revenue gains excluding the effect of adding $H3L$ SKUs.

The retailer decided to test a portion of our recommendations by adding eleven of the 47 SKUs we had recommended be added to the assortment, ten $H3L$ SKUs and one $H1H$. Given the lead time involved in procuring these new tires, the changes to the assortment were implemented only in July 2005. The retailer used the same assortment in all stores.

The validation analysis that we conducted earlier used data for the period January 2005 to July 2005 during which none of the eleven new SKUs were included in the assortment. Hence, our first objective was to test the performance of our demand estimation procedure for forecasting sales of the eleven new tires introduced into the assortment in the period July 2005 to December 2005. To achieve this, we first re-calibrated the model by using sales data from January 2005 to June 2005 to estimate demand and substitution parameters available immediately prior to the July to December 2005 implementation period. We then used the revised demand estimates to forecast sales shares, at each store, for the newly introduced SKUs, for the period July 2005 to December 2005. Table 10 shows a comparison of the predicted vs. actual chain sales shares for the new SKUs. The sales-weighted MAD across all SKUs is 17%, and the MAPE is 19.1%, which compares favorably to the 30.7% MAPE for chain sales of two new SKUs reported by Fader and Hardie (1996).

Our second objective was to estimate the revenue lift that the retailer achieved by implementing a portion of our recommendations. Estimating the change in revenue from the current to the implemented assortment was complicated because, in addition to adding eleven new SKUs to the assortment, the retailer deleted more than eleven 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 $N_s - 11$ 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 pre implementation assortment at each store and that 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 11 gives revenue estimates and percentage improvement over the current baseline assortment for these periods.

Note that in the validation period, we would have estimated a 13.1% revenue increase (from \$74.9 million to \$84.7 million) from the implemented assortment, whereas the actual increase was 5.8% (from \$72.3 million to \$76.5 million) due to change in parameters over time. We note that a 5.8% improvement is large relative to what retailers typically achieve through enhancements to existing stores. 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 (Canadian Tire Corporation Limited 2007).

Similar to the analysis conducted for Snack Cakes, we find that the higher MAD for the validation and implementation periods can be explained partly by sales trends that cause the parameter values to change. For example, the aging over time of the car models that use a particular tire impacts the demand for that tire. The demand for a tire type initially increases as cars that use that tire age and need replacement tires, but eventually declines as those cars becomes old enough that they begin to exit the population. Moreover, the retailer changed relative prices of the six brand-warranties from the calibration to the demand periods, which impacted the six demand shares. Table 12 shows how changes in relative prices across brand-warranties relates to systematic changes in their demand shares over time, which is clearly not captured in our current model. In particular, it is quite striking in this Table that 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%.

5.3 Major Auto Aftermarket Parts Retailer

This 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.

We worked with the appearance chemicals category manager and other members of her team, as well as a staff team that supported category management. On June 1 we received store-SKU sales history on the 160 SKUs currently carried in this category at 3236 stores for the period May 1, 2008 to April 30, 2009. We applied the methodology described in this paper and reported final results to the senior management of the retailer on July 25. These recommendations were accepted for implementation with the few modifications described below. The new assortment was implemented during 1/17/2010 – 1/23/2010 and subsequent sales results have been tracked. We describe below the details of this application.

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) 9 brands and (3) 3 quality levels, denoted as one of the three levels 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 didn't 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 x 17 matrix of possible attribute levels. Of the 45 x 17 − 130 = 635 attribute combinations not carried by the retailer, only 24 were available in the market.

We applied the methodology described in this paper to estimate demand for available SKUs not carried. No substitution parameters were used in the model for the following reason. If a brand-quality level were 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 this data was not available. The Mean Absolute Deviation of forecasts of existing SKUs across all 3,236 stores was 20.2%.

After parameter estimation, we applied the greedy heuristic to the problem of choosing 130 out of a potential 154 attribute combinations so as to maximize predicted revenue. 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, five being the greatest number of assortments the retailer specified they would consider, given the administrative load of multiple assortments. The estimation process took approximately 5 seconds per store on an Intel Core 2 Duo 2 GHz processor. The computation time for the greedy algorithm to generate five different assortments for all stores on the same computer was approximately 75 minutes.

Table 13 shows the estimated revenue increase 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. A single assortment resulted in a 11.8% increase in revenues, while store specific assortments lead to a 14.2% increase in revenues, implying a localization lift of 2.4%. The two cluster solution was selected for implementation. In the revised assortments, 20 SKUs in cluster 1 with the lowest estimated revenue were replaced by 20 new SKUs. In cluster 2, 19 existing SKUs were replaced.

Table 14 shows the distribution of revenue for the top 24 segment/sizes by segment/size, brand and quality level for the two clusters and some demographic data, including the percentage of people in the categories 'suburban' or 'urban/bilingual' in the zip codes in which the stores of each cluster are located. Noteworthy differences for cluster 2 are a higher demand for tire related products, higher demand for brand 2, lower demand for brand 5 and a higher percentage of urban/bilingual. Table 15 shows these same data for the two stores with highest and lowest percentage suburban. The differences noted above persist, and to a much great degree.

The retailer largely adopted our recommendations. They 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 one and twenty five to cluster two. The sales-weighted MAD across the new SKUs added is 26.8%, and the MAPE is 28.7%, which again compares favorably to the 30.7% MAPE for chain sales of two new SKUs reported by Fader and Hardie (1996). 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 one and twenty three from cluster two. 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.

With respect to localization, the retailer regarded cluster 1 as a base case that was representative of the chain as a whole and cluster two as a subset of stores differentiated by the higher level of tire related purchases, a preference for brand two, and a higher percentage of 'urban/bilingual'. This store segmentation was compelling for the retailer and agreed with more qualitative market research inputs they had received. 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, 2010 – 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, 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 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 SKUs deleted. 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 vs. seventeen deleted and twenty five SKUs were added to cluster 2 vs. twenty three 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 twenty two and twenty five SKUs for clusters one and two, respectively, with lowest revenue in the 2009 evaluation period in estimating deletion revenue, thereby equalizing the add and delete counts. Revenue is affected by a variety of factors other then assortment, including weather, the economy and competitive activity. We measured the impact of these other factors by the ratio of 2010 to 2009 revenue for kept SKUs and estimated the 2010 revenue of the deleted SKUs as their 2009 revenue times this factor.

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 7 weeks to achieve a steady steady 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 two stores that we were not able to measure. The retailer's appearance chemicals team agreed with our assessment of benefits and believed that the re-assortment exercise had been a success.

6 Analysis of Results

6.1 Understanding Localization Revenue Lift

The Localization Lift, defined as the revenue increase from using store specific assortments vs. a single assortment for the chain was 12.2%, 5.8% and 2.4% for the snack cakes, tires and appearance chemicals examples described in the previous sections. As we sought to understand what features of the problem data cause these differences in Localization Lift, our first thought was that Localization Lift must be driven by demand variation across stores. We thus calculated a coefficient of demand variation (COV) 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 of revenue shares of SKU i across all stores. COV for snack cakes tires and appearance chemicals was 17%, 11% and 10% respectively. These values shed some light on variation in Localization Lift in that the highest COV matches the highest lift, for snack cakes, but leave open the drivers of variation in Localization Lift between tires and appearance chemicals, where the Localization Lift varies by a factor of three while the COV's are nearly equal.

To better understand this issue, we examined the data more closely and made two observations. First, SKUs can be segmented 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. While there may be substantial variation in demand across stores for the first groups, none of this variation impacts Localization Lift.

For example, in the case of appearance chemicals, we see in Table 14 that the best selling segment-size for the chain is Tire Dressings/Shines TRIGGER L. Table 15 shows substantial difference in the sales rate between two stores for Tire Dressings/Shines TRIGGER L. The best selling single SKU in this segment-size, as defined by brand and quality level, accounted for 5.5% of revenue in Store A and 16.6% in Store B, a 3.3 to 1 difference. Yet even though the SKU sold much worse in Store A, with a revenue share of 5.5%, it clearly made sense to have this SKU in a revenue maximizing assortment for Store A, and hence this difference in sales rate had no impact on the Localization Lift.

Secondly, we noticed substantial variation in the breadth of assortment carried, from 40 out of 92 possible SKUs for snack cakes to 130 out 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.

The example in Table 16 is designed to illustrate how these patterns can occur. The example compares COV and Localization Lift for two demand cases, and within each case, for K equal to 3 or 4. The price of all SKUs is \$1, so unit demand and revenue are the same. Moreover, total demand is equal for the two stores, so there is no distinction between demand units and demand share. We also assume there is no substitution. The variance column gives the variance in demand across the two stores for each SKU. COV, as we have defined above is the square root of total variance divided by total demand.

In all cases the chain optimal assortment is SKUs 1 through K. For $K = 3$, the optimal assortment for store 1 is SKUs 1, 2 and 3 and for store 2, SKUs 1, 2 and 4. For $K = 4$, SKUs 1 to 4 are an optimal assortment for both stores.

Considering the case $K = 3$, note that Case 1 has the highest COV but the lowest Localization Lift. This happens because almost all of the inter store demand variation occurs for SKUs that are in either both store optimal assortments or neither and hence none of this variation impacts Localization Lift. By contrast, in Case 2, all of the inter store demand variation occurs for SKUs 3 and 4, the very SKUs that differ in the store optimal assortments.

Note also that when K=4, the lift in both cases drops to 0, demonstrating the impact that breath of assortment can have on lift.

Motivated by what we saw in the data for the three applications and by the features demonstrated by the example in Table 16, we defined two additional metrics which we hypothesize would be correlated with Localization Lift. COV Select (K) is the coefficient of variation for SKUs which are in some but not all store optimal assortments. This metric is a function of K because store optimal assortments depend on K. We re-label our original coefficient of variation metric as COV – All to emphasize its difference with COV – Select (K).

We also define Chain Optimal Share (K) to be the share of the total potential revenue $\sum_{s \in M} \sum_{i \in N} p_i D_i^s$ achieved by a single, chain optimal assortment with K SKUs. So far we haven't 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.

Table 17 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 (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 the high value of Chain Optimal Share (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 Percent 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.

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 18 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 against the three dependent variables; the results are summarized in Table 18. The regression results confirm our hypothesis that Localization Lift depends mainly on Chain Optimal Share and COV - Select, which are highly significant, and not on COV - All, which is insignificant.

6.2 Performance of Heuristics

We consider the quality of the solutions produced by the greedy and interchange heuristics. Both of these heuristics have been used previously for assortment optimization, greedy in Kök and Fisher (2007) and interchange in Chong et al. (2001). Belloni et al. (2008) computationally evaluate greedy and interchange for product line design, a problem similar in structure to assortment optimization, and find that on real problems greedy achieves 98.4% of maximum profit on average and greedy followed by interchange, 99.9%.

Note that if there is no substitution, greedy finds an optimal solution since the assortment problem is then maximization of a linear function subject to an upper bound on the sum of the variables. Thus greedy found optimal solutions for appearance chemicals where there was no substitution, so we will restrict our attention here to the snack cakes and tires applications.

Note that Table 3 shows an interesting property for the assortment for a typical snack cakes store. The flavors are sorted by demand rank and as we move down the list of flavors, the number of brand-sizes offered steadily decreases. The following Theorem will establish conditions under which this property always holds for an optimal assortment. We'll use the theorem to generate a sample of large problems with known optimal solution against which we can test the effectiveness of the greedy and interchange heuristics.

Theorem 1. Consider an assortment planning problem for a store s with the following characteristics.

1. $A = 2$
2. $\pi_{1uv} = 0$ for all u and v, i.e., no substitution across levels of attribute 1
3. If $p(u, v)$ denotes the price of a SKU with levels u and v for attributes 1 and 2 respectively, then there exists constants p_{1u} and p_{2v}, $u = 1, 2, \ldots N_1$ and $v = 1, 2, \ldots, N_2$ such that $p(u, v) = p_{1u}p_{2v}$.
4. The levels of attribute 1 are indexed so that $p_{1u}f_{1u} \geq p_{1,u+1}f_{1,u+1}$, $u = 1, 2, \ldots, N_1 - 1$.

Given an assortment S, let $S(u)$ denote the set of levels of attribute 2 that are present in the SKUs in S that have level u of attribute 1, where $S(u)$ is the empty set if there is no SKU in S with level u of attribute 1.

Then there exists an optimal assortment satisfying $\mid S(u) \mid \geq \mid S(u+1) \mid$, $u = 1, 2, \ldots, N_1 - 1$.

Proof. Let D denote total unit demand at the store, $k(u) = \mid S(u) \mid$, and for $v \notin S(u)$, $w(v, S(u)) = \arg\max_{w \in S(u)} \pi_{2vw}$. $w(v, S(u))$ is the level of attribute two that is the best substitute in the assortment S for a customer desiring level v of attribute 2. By the assumption of the theorem, we can write the assortment optimization problem for store s as

$$\max D \sum_{u=1}^{N_1} p_{1u} f_{1u} \max_{S(u), |S(u)|=k(u)} \left[\sum_{v \in S(u)} p_{2v} f_{2v} + \sum_{v \notin S(u)} p_{2w(v,S(u))} f_{2v} \pi_{2vw(v,S(u))} \right]$$

Note that the expression in square brackets maximized in the choice of $S(u)$ does not depend on u. So letting $Z(k) = \max_{S(u),|S(u)|=k(u)} \left[\sum_{v\in S(u)} p_{2v}f_{2v} + \sum_{v\notin S(u)} p_{2w(v,S(u))}f_{2v}\pi_{2vw(v,S(u))}\right]$, we can express assortment revenue as $D\sum_{u=1}^{N_1} p_{1u}f_{1u}Z(k(u))$.

If the theorem is violated, then there is a u such that $k(u) < k(u+1)$. But then $Z(k(u)) \leq Z(k(u+1))$, because $Z(k(u+1))$ is the optimal value of a maximization problem that is less constrained than that which determines $Z(k(u))$. This, together with assumption (4) of the theorem implies

$$p_{1u}f_{1u}Z(k(u+1)) + p_{1,u+1}f_{1,u+1}Z(k(u)) \geq p_{1u}f_{1u}Z(k(u)) + p_{1,u+1}f_{1,u+1}Z(k(u+1)).$$

So we can revise the solution by assigning the set $S(k(u+1))$ for attribute level u and the set $S(k(u))$ for attribute level $u+1$ without reducing revenue. The revision removes this violation of the theorem. Repeated application of this step will produce an optimal solution satisfying the condition of the theorem.

The snack cakes and tires applications satisfy conditions 1 and 2 of the theorem. With prices as determined by the hedonic regression, they also satisfy 3, and 4 is easily satisfied by indexing attribute 1 as required. In the applications, we used real, not estimated, prices for current SKUs, so the problems as we solved them did not exactly satisfy the conditions of the theorem. However, we subsequently used the theorem to find optimal solutions to store level assortment problems for all stores for the cakes and tires applications with estimated prices and then tested the heuristics against these known optima.

To find the optimal solutions using the theorem, we first note that because the expressions in square brackets in the proof doesn't depend on the level u of the first attribute, the same set is optimal for all u with a common value of $k(u)$. Moreover, the number of levels for the second attribute (4 for cakes, 6 for tires) is small enough that we can enumerate all subsets of these attribute levels, and for each possible cardinality k, identify the optimal subset and associated values $Z(k)$. Then the assortment optimization problem reduces to finding optimal values of $k(u)$, $u = 1, 2, \ldots, N_1$, satisfying $k(u) \geq k(u+1)$ and $\sum_{u=1}^{N_1} k(u) = K$. The number of feasible values of $k(u)$ is small enough to allow for complete enumeration.

Using this approach we found optimal assortments with estimated prices for each of the 140 stores in the snack cakes application and the 574 stores in the tires applications. This gave us a sample of 714 problems of realistic size and known optimal assortments on which to test the performance of our heuristics. The results of applying greedy to all of these problems are reported in Table 19 and show that greedy finds near optimal solutions and performs similarly on our applications to what Belloni et al. (2008) found for the product line design problem.

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.

7 Conclusions

We have formulated a process for finding optimal assortments, comprised of a demand model, and estimation approach and heuristics for choosing assortments. We have applied this process to real data from three applications and shown that the approach produces accurate forecasts for new SKUs. Our recommendations were implemented in two of the cases. We measured the impact based on actual sales and found the assortment revisions had produced revenue increases of 5.8% and 3.6%, which 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 was adequate to achieve significant benefits in implementation. The only prior reported results for forecasts of new retail SKUs is Fader and Hardie (1996), who reported an average MAPE of 30.7% for two new grocery SKUs. We found a MAPE of 16.2% for one new snack cakes SKU, 19.1% for 11 new tires SKUs, and 28.7% for 25 new appearance chemical SKUs, somewhat improving on the results of Fader and Hardie (1996). Nonetheless, the errors were great enough to reduce the revenue increase by about half from the fit to validation samples, so improving forecast accuracy would be a useful focus of future research.

2. Sales is not true demand, but demand distorted by the assortment offered. We don't 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 60% 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 have a major impact on the optimal assortment. In the snack cakes 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 family size of the optimized assortment.

4. We were able to use demographic data to confirm our parameter estimates in the tire and appearance chemicals examples. In particular, 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 showed that this difference is not driven by demand variation across stores, but by the percent of maximum revenue captured by a chain optimal assortment and by variation in demand for those SKUs that vary across store optimal assortments.

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

1 assortment to 140 assortments.

7. There may be interaction between attribute levels not captured by our simplest 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.

8 Tables

Table 1: Snack Cakes: Demand Share Estimates (Averaged across Stores)

Brand Size	**Sales Share**	**Demand Share**
B_1S	67%	61%
B_2S	24%	27%
B_1F	4%	6%
B_2F	5%	6%

Table 2: Snack Cakes: Substitution Probability Estimates (Averaged across Stores)

	B_1S	B_2S	B_1F	B_2F
B_1S	1	18%	0	0
B_2S	26%	1	0	0
B_1F	0	0	1	89%
B_2F	0	0	22%	1

Table 3: 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

Table 4: Management's estimate of the most likely substitution probabilities

From	**To** NH	NM	H1H	H2H	H2M	H3L
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

S =Somewhat Likely, L =Likely, M = Most Likely

Table 5: Snack Cakes: Number of Brand-Sizes by Flavor for the Optimal Assortment at a Representative Store

Flavor	**Demand Rank**	**Number of Brand Sizes in the Optimal Assortment**
Cinnamon	1	3
Chocolate	2	3
Peanut Butter	3	3
Butterscotch	4	3
Butter	5	3
Vanilla	6	3
Raspberry	7	3
Fudge	8	3
Honey	9	3
Buttercream	10	3
Choc. Chip	11	2
Cherry/Cheese	12	2
Cheese	13	2
Coconut	14	1
Oatmeal/Raisin	15	1
Jelly	16	1
Vanilla/Chocolate	17	1
Pineapple/Cheese	18	0
Blueberry	19	0
Marshmallow	20	0
Apple	21	0
Cream	22	0
Glazed	23	0

Table 6: 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%

Table 7: Tires: Substitution Probability Estimates (Averaged across Stores)

Substitution Likelihood	**Probability**
Somewhat Likely	2%
Likely	6%
Most Likely	45%

Table 8: Tires: Assortment Changes for a Representative Store

Brand Warranty	**SKUs Added**	**SKUs Deleted**
NH	11	1
NM		5
H1H	1	8
H2H	19	
H2M	2	32
H3L	14	1
Total	47	47

Table 9: Tires: Revenue Lift vs. π_M

π_M	**Revenues ($ million)**	**Increase (percent)**
0.00	337.7	322.1
0.23	133.1	66.4
0.45	104.1	30.1
0.73	95.2	19.0
1.00	87.2	9.0

Table 10: Tires: Actual vs. Predicted Chain Level Sales of New SKUs

		Sales Share (%)		
Size	**Brand**	**Actual**	**Predicted**	**% Error**
P225/60R16	H3L	3.0	3.4	10.9
P215/70R15	H3L	3.1	3.5	12.6
P205/65R15	H3L	3.2	4.4	37.1
P205/70R15	H3L	2.8	2.7	5.1
P195/65R15	H3L	2.2	2.4	7.6
P215/65R15	H3L	1.0	1.2	25.3
P205/55R16	H3L	1.2	1.1	6.2
P215/60R16	H3L	1.1	1.3	19.4
P215/70R14	H3L	0.9	1.2	37.2
195/70R14	H2H	0.8	1.0	29.5

The sales-weighted MAD across all new SKUs is 17%

Table 11: Tires: Revenue Estimates for Current, Implemented and Optimized Assortments (percentage improvement over current revenues given in parenthesis)

Revenues ($$ million)	**Parameter Estimates Used**		
Assortment	**Jul 04 - Dec 04**	**Jan 05 - Jun 05**	**Jul 05 - Dec 05**
Current Assortment (Jul 04 - Dec 04)	80.2	74.9	72.3
Modified Implementation (Jul 05 - Dec 05)	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)

Table 12: 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

Table 13: Estimated Revenue Increases (%) vs. Number of Store Clusters

No. of Clusters	2183 Stores (warm up)	3236 Stores (implementation)
1	14.1	11.8
2	14.4	11.9
3	14.6	12.0
5	14.9	

Table 14: Cluster Statistics for Top 25 SKUs for Appearance Chemicals

Revenue Distribution by Product type - package size, Quality and Brand

Segment - package size	**Cluster 1**	**Cluster 2**	**Total**
Tire Dressings / Shines TRIGGER L	7.1%	12.6%	8.8%
Multi-Purpose Protectants S	7.2%	8.4%	7.6%
Washes L	7.4%	6.6%	7.1%
Tire Dressings / Shines AEROSOL S	5.0%	7.4%	5.8%
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%
Washes S	4.9%	3.8%	4.6%
Multi-Purpose Protectants L	4.0%	3.8%	4.0%
Tire Foams (multi-purpose) AEROSOL S	3.1%	4.1%	3.4%
Carpet / Upholstery Cleaners AEROSOL S	3.2%	3.0%	3.1%
Spray Wax S	3.0%	2.6%	2.9%
Paste Wax S	3.0%	2.2%	2.8%
Wheel Care / Cleaner. ALL TRIGGER S	2.7%	2.8%	2.7%
Spray Detailers L	3.0%	2.1%	2.7%
Leather Cleaners / Conditioners TRIGGER S	2.7%	2.0%	2.5%
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%
Scratch Removers S	1.8%	1.1%	1.6%
Rubbing / Polishing Compounds L	1.9%	0.9%	1.6%
Glass Cleaners TRIGGER L	1.5%	1.6%	1.5%
Plastic / Lens Cleaners, Polishes & Repair S	1.6%	1.3%	1.5%
Glass Cleaners AEROSOL S	1.6%	1.4%	1.5%

Quality	**Cluster 1**	**Cluster 2**	**Total**
Good	12%	14%	13%
Better	60%	67%	62%
Best	28%	20%	25%
Total	**100%**	**100%**	**100%**

Brand	**Cluster 1**	**Cluster 2**	**Total**
Brand 1	18%	20%	19%
Brand 2	7%	11%	8%
Brand 3	6%	6%	6%
Brand 4	19%	16%	18%
Brand 5	12%	7%	10%
Brand 6	6%	7%	7%
Brand 7	2%	2%	2%
Brand 8	11%	10%	11%
Brand 9	19%	21%	19%
Total	**100%**	**100%**	**100%**

Demographic statistics

	Cluster 1	**Cluster 2**	**Total**
Income Index	0.95	0.80	0.90
% Suburban	85%	62%	78%
% Urban/bilingual	16%	42%	24%
Store count	2263	973	15%

Table 15: Statistics for Stores with Maximum and Minimum Percent Suburban

Revenue Distribution by Product type - package size, Quality and Brand

Segment - package size	Store A	Store B
Tire Dressings / Shines TRIGGER L	5.5%	16.6%
Multi-Purpose Protectants S	5.3%	10.6%
Washes L	7.8%	8.5%
Tire Dressings / Shines AEROSOL S	4.6%	8.8%
Tire Cleaners TRIGGER L	6.5%	5.3%
Liquid Wax S	8.1%	2.3%
Wash and Wax S	2.8%	5.0%
Washes S	4.5%	3.4%
Multi-Purpose Protectants L	4.8%	3.2%
Tire Foams (multi-purpose) AEROSOL S	3.4%	3.0%
Carpet / Upholstery Cleaners AEROSOL S	2.3%	3.2%
Spray Wax S	4.1%	2.0%
Paste Wax S	2.2%	3.0%
Wheel Care / Cleaner. ALL TRIGGER S	2.3%	2.8%
Spray Detailers L	5.1%	1.2%
Leather Cleaners / Conditioners TRIGGER S	3.2%	1.2%
Wheel Care / Cleaner. ALL TRIGGER L	3.4%	1.0%
Rubbing / Polishing Compounds S	0.5%	2.5%
Tire Dressings / Shines BOTTLE GEL S	1.5%	1.8%
Scratch Removers S	1.4%	1.6%
Rubbing / Polishing Compounds L	2.0%	1.1%
Glass Cleaners TRIGGER L	2.0%	1.1%
Plastic / Lens Cleaners, Polishes & Repair S	1.5%	1.3%

Quality	Store A	Store B
Good	13%	18%
Better	57%	66%
Best	31%	16%
Total	**100%**	**100%**

Brand	Store A	Store B
Brand 1	16%	18%
Brand 2	6%	15%
Brand 3	5%	6%
Brand 4	16%	12%
Brand 5	17%	4%
Brand 6	7%	10%
Brand 7	2%	2%
Brand 8	10%	10%
Brand 9	22%	23%
Total	**100%**	**100%**

Demographic statistics

	Store A	Store B
Income Index	0.98	0.88
% Suburban	98%	8%
% Urban/bilingual	1%	91%
Location	Marietta, OH	Lithonia, GA

Table 16: Example to Illustrate Drivers of Localization

	Demand			Demand		
SKU	**Store 1**	**Store 2**	**Variance**	**Store 1**	**Store 2**	**Variance**
1	50	100	1250	75	75	0
2	100	50	1250	75	75	0
3	40	50	50	50	0	1250
4	50	40	50	0	50	1250
5	40	0	800	25	25	0
6	0	40	800	25	25	0
7	40	0	800	25	25	0
8	0	40	800	25	25	0
9	40	0	800	25	25	0
10	0	40	800	25	25	0
11	40	0	800	25	25	0
12	0	40	800	25	25	0
Totals	400	400	9000	400	400	2500
COV			11.9%			6.4%

K	**Chain Optimal**	**Localized Optimal**	**Localization Lift**	**Chain Optimal**	**Localized Optimal**	**Localization Lift**
3	390	400	2.6%	350	400	14.3%
4	480	480	0	400	400	

Table 17: Explaining Localization Lift

Category	**Lift**	**COV - All**	**COV - Select**	**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

Table 18: 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			
Minimum K	10	10	5
Maximum K	40	100	40
Minimum Stores	10	20	80
Maximum Stores	30	60	240
# of Instances Simulated	100	100	100

Table 19: Performance of Greedy Heuristic

Product	N_1	N_2	Average Performance (%)	Finds Optimal Solution (%)	Average Time per Trial (seconds)
Cakes	23	4	97.2	74.6	0.14
Tires	64	6	98.5	80.3	0.34

9 Figures

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

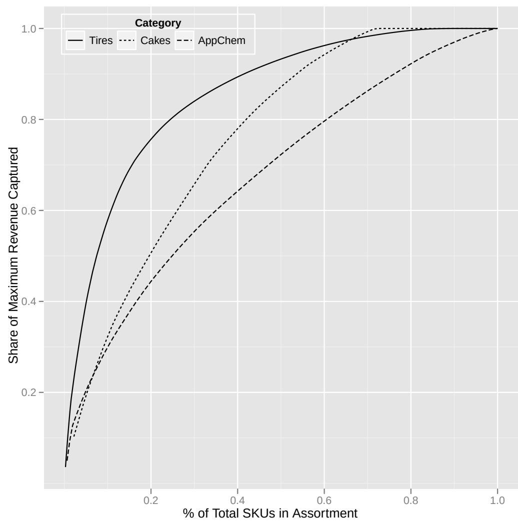


Figure 2: Share of $H3L$ ($H2H, H2M$) is Negatively (Positively) Correlated with Income

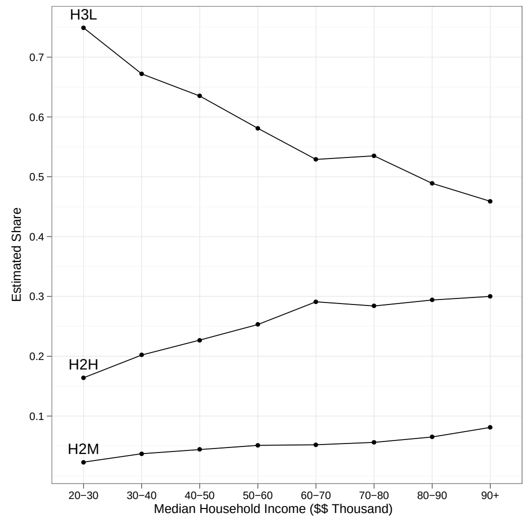
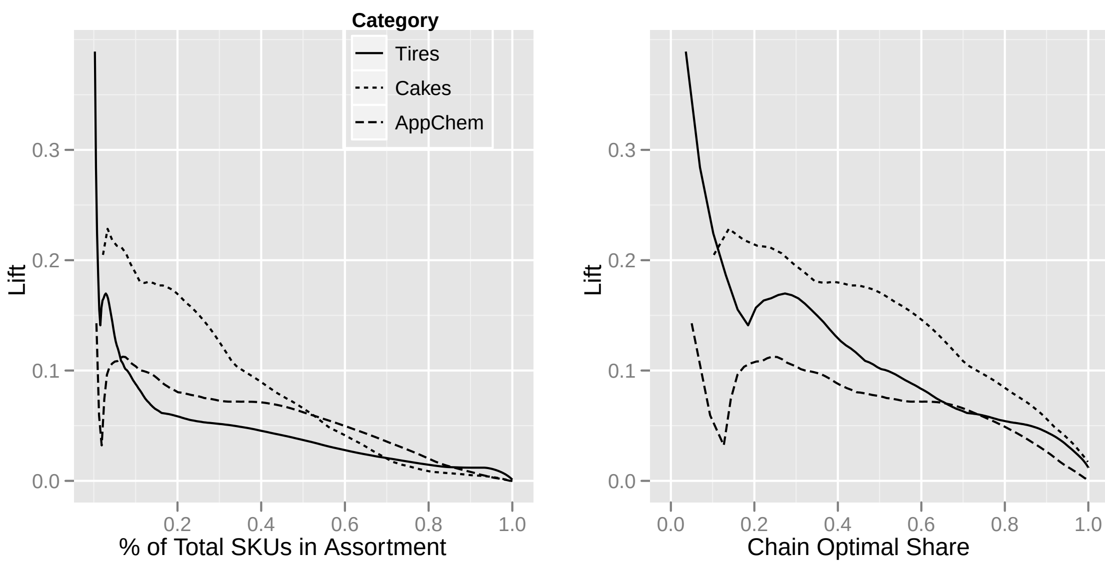


Figure 3: Localization Lift versus Maximum Total SKUs in the Assortment and Chain Optimal Share



References

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

Bell, D.R., A. Bonfrer, P.K. Chintagunta. 2005. Recovering SKU-level preferences and response sensitivities from market share models estimated on item aggregates. *Journal of Marketing Research* **42**(2) 169–182.

Belloni, A., R. Freund, M. Selove, D. Simester. 2008. Optimizing Product Line Designs: Efficient Methods and Comparisons. *Management Science* **54**(9) 1544.

Beswick, P., M. Isotta. 2010. Making the right choices: SKU rationalization in retail. Oliver Wyman Working Paper.

Cachon, G., C. Terwiesch, Y. Xu. 2005. Retail assortment planning in the presence of consumer search. *Manufacturing & Service Operations Management* **7**(4) 330.

Canadian Tire Corporation Limited. 2007. Annual report.

Caro, F., J. Gallien. 2007. Dynamic assortment with demand learning for seasonal consumer goods. *Management Science* **53**(2) 276–276.

Chong, J. K., T. H. Ho, C. S. Tang. 2001. A modeling framework for category assortment planning. *Manufacturing & Service Operations Management* **3**(3) 191–210.

Dempster, AP, NM Laird, DB Rubin. 1977. Maximum likelihood from incomplete data via the em algorithm. *Journal of the Royal Statistical Society. Series B (Methodological)* **39**(1) 1–38.

Dobson, G., S. Kalish. 1988. Positioning and pricing a product line. *Marketing Science* **7**(2) 107–125.

Fader, Peter S., Bruce G. S. Hardie. 1996. Modeling consumer choice among skus. *Journal of Marketing Research* **33**(4) 442–452.

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

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

Goodman, A. C. 1998. Andrew court and the invention of hedonic price analysis. *Journal of Urban Economics* **44**(2) 291–298.

Green, P.E., A.M. Krieger. 1985. Models and heuristics for product line selection. *Marketing Science* **4**(1) 1–19.

Green, P.E., A.M. Krieger. 1987a. A consumer-based approach to designing product line extensions. *Journal of Product Innovation Management* **4**(1) 21–32.

Green, P.E., A.M. Krieger. 1987b. A simple heuristic for selecting 'good' products in conjoint analysis. *Applications of Management Science* **5** 131–153.

Green, P.E., A.M. Krieger. 1992. An application of a product positioning model to pharmaceutical products. *Marketing Science* **11**(2) 117–132.

Guadagni, P.M., J.D.C. Little. 1983. A logit model of brand choice calibrated on scanner data. *Marketing Science* **2**(3) 203–238.

Hotelling, H. 1929. Stability in competitition. *Economic Journal* **39** 41–57.

Kamakura, W.A., G.J. Russell. 1989. A probabilistic choice model for market segmentation and elasticity structure. *Journal of Marketing Research* **26**(4) 379–390.

Kohli, R., R. Sukumar. 1990. Heuristics for Product-Line Design Using Conjoint Analysis. *Management Science* **36**(12) 1464–1478.

Kök, G., M. L. Fisher. 2007. Demand Estimation and Assortment Optimization under Substitution: Methodology and Application. *Operations Research* **55**(6) 1001–1021.

Kök, G., M. L. Fisher, R. Vaidyanathan. 2008. Assortment planning: Review of literature and industrial practice. Invited chapter in Retail Supply Chain Management, Eds. N. Agrawal, S.A. Smith. Kluwer Publishers.

Lancaster, K. 1975. Socially Optimal Product Differentiation. *The American Economic Review* **65**(4) 567–585.

Lancaster, K. J. 1966. A new approach to consumer behavior theory. *Journal of Political Economy* **74** 132–157.

Maddah, B., EK Bish. 2007. Joint pricing, assortment, and inventory decisions for a retailer's product line. *Naval Research Logistics* **54**(3) 315.

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

McBride, R.D., F.S. Zufryden. 1988. An integer programming approach to the optimal product line selection problem. *Marketing Science* **7** 126–140.

McGregor, Jena. May 15, 2008. At Best Buy, marketing goes micro. *BusinessWeek* .

O'Connell, Vanessa. April 21, 2008. Reversing field, Macy's goes local. *Wall Street Journal* .

Pakes, Ariel. 2003. A reconsideration of hedonic price indexes with an application to PC's. *The American Economic Review* **93**(5) 1578–1596.

Rosen, S. 1974. Hedonic prices and implicit markets: Product differentiation in pure competition. *The Journal of Political Economy* **82**(1) 34–55.

Smith, S. A., N. Agrawal. 2000. Management of multi-item retail inventory systems with demand substitution. *Operations Research* **48**(1) 50–64.

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

Vulcano, G., G. van Ryzin, R. Ratliff. 2009. Estimating primary demand for substitutable products from sales transaction data. Working Paper, NYU Stern.

Zimmerman, A. September 7, 2006. To boost sales, Wal-Mart drops one-size-fits-all approach. *Wall Sreet Journal Online* A1.

Zimmerman, Ann. October 7, 2008. Home Depot learns to go local. *Wall Street Journal* .