Chapter 10
Managing Variety on the Retail Shelf: Using Household Scanner Panel Data to Rationalize Assortments

Ravi Anupindi, Sachin Gupta, and M.A. Venkataramanan

1 Introduction

Two fundamental retailer decisions are which items to stock in a category (the assortment decision) and how much to stock of each item (the inventory decision). While these decisions have always been key to retailer profitability, they have received renewed attention because of industry initiatives labeled Efficient Consumer Response (ECR). Category Management, a component of ECR, emphasizes the need to recognize the inter-relatedness (e.g., substitutability) of items within a category when making decisions. Thus, categories need to be managed as strategic business units, with an emphasis on total category performance. Point-of-sale information can potentially play a critical role in providing insights into consumer behavior to help develop sound category strategies.

Retailers recognize that wider assortments help their business by catering to the needs of multiple consumer segments (Coughlan et al. 2006), as well as by offering variety to variety-seeking consumers. However, there are limits to the value of variety. Adding items with small differences offers little in the way of "real" variety to the consumer (Boatwright and Nunes 2001), yet adds to costs of operations such as

R. Anupindi (✉)
David B Hermelin Professor of Business Administration, Professor of Technology and Operations, Stephen M. Ross School of Business, University of Michigan, Ann Arbor, MI, USA
e-mail: anupindi@umich.edu

S. Gupta
Director, Graduate Studies and Henrietta Johnson Louis Professor of Management, Johnson Graduate School of Management, Cornell University, Ithaca, NY, USA

M.A. Venkataramanan
Vice Provost for Strategic Initiatives, Jack R. Wentworth Professor, School of Business, Indiana University, Bloomington, IN, USA

N. Agrawal, S.A. Smith (eds.), *Retail Supply Chain Management*,
International Series in Operations Research & Management Science 223,
DOI 10.1007/978-1-4899-7562-1_10

administrative costs and cost of warehouse space. The sharp growth of warehouse clubs and deep discount drug stores in recent years is attributed, in part, to their cost advantages arising from their limited variety offering. The resultant loss of market share has re-focussed attention of supermarkets on the need to manage variety. It is believed that there is substantial potential for lowering supermarket operating costs without hurting business by making store assortments more efficient; see, for example, a report by the Food Marketing Institute (1993).

Managing retail space entails solving two types of problems. The first is allocating space to categories, called the *inter-category* space allocation problem. The second is allocating space to items within a category or the *intra-category* space allocation problem. This second problem is often referred to as the assortment problem. Ideally, assortment decisions need to incorporate a variety of factors. On the demand side, one needs to consider the (heterogenous) customer purchase behavior including substitution patterns when their preferred items are not available (either temporarily due to stock-out or permanently due to limited assortment), the stochastic nature of demand arising due to the uncertainty inherent in consumer choice, the effect of product display on sales, etc. On the supply side, retailers face a finite shelf-space constraint for a category and incur fixed costs to include items in the assortment. Further, since limited assortments may have longer term consequences on profitability, a retailer needs to balance current profits with implications of the assortment on future profits. Finally, such a model for decision making should be driven by actual data and the solution strategy should be scalable to address the large problem sizes that any realistic assortment decision would entail.

In this chapter, we outline a modeling framework that incorporates some of the above features to assist the retailer in determining the optimal subset of items to carry in a category, from the set currently carried, and the quantity to stock of each item. We propose the use of household purchase data collected via scanners to estimate intrinsic preferences of consumers and to infer their substitution patterns. Such information is key to ensuring that the assortment carried caters to heterogeneous consumers' tastes, while avoiding unnecessary and expensive duplication.

Previous research on the retailer's assortment problem has typically not modeled consumer substitution behavior explicitly. Empirical evidence from several studies suggests that in packaged goods markets, consumers are often willing to substitute a less preferred item for their (non-available) preferred item. A Food Marketing Institute survey reports that only 12–18 % of shoppers said they would not buy an item on a shopping trip if their favorite brand-size was not available; the rest indicated they would be willing to buy another size of the same brand, or switch brands. A number of other studies (Emmelhainz et al. 1991; Carpenter and Lehmann 1985; Urban et al. 1984; Gruen et al. 2002) support a similar conclusion. A 1993 study by Willard Bishop Consulting Ltd. and Information Resources, Inc. found that when duplicative items were removed, 80 % of consumers saw no difference (Business Week 1996). Other evidence suggests that consumers make about two-thirds of their purchase decisions about grocery and health-and-beauty products while they are in the store (Nielsen Marketing Research 1992). Thus, it is important to take account of substitution behavior of consumers when rationalizing assortments.

It is likely that consumers who do not find their preferred item in the store assortment are not fully satisfied, whether or not they buy another item in the category. The decision to rationalize assortments needs to take account of the potential adverse impact on customer retention. Traditional formulations of the assortment problem typically assume that the retailer is a myopic profit maximizer. Such formulations disregard the longer-term adverse impact on profits of not satisfying consumers' demand for their preferred items. In our proposed formulation, the retailer's objective function is a weighted sum of profits and a penalty for disutility caused to consumers who do not find their preferred items in the assortment. The rationale for including a penalty is that dissatisfied customers may take their future business elsewhere, thereby hurting longer term profits, even if they purchase less preferred items in the current period. Our proposed model can be used by a retailer to balance short term profits and customer disutility when choosing assortments.

Another contrast of our proposed approach with previous research lies in our accommodation of differences in item preferences between consumers. Most previous work assumes an aggregate demand model. Aggregate demand specifications do not allow us to distinguish between the extent of disutility or dissatisfaction caused by not stocking a particular item to, for example, more versus less loyal groups of consumers. Clearly this distinction is relevant for a retailer who cares about retaining customers in the longer run. The existence of consumer heterogeneity has been established by a number of previous empirical studies. Our proposed model allows for completely idiosyncratic patterns of substitution, as well as disutility due to non-stocking, between consumers.

To demonstrate an empirical application of the proposed model, we estimate consumer preferences for eight items in the canned tuna category using household scanner panel data, a commonly available source of market research information. A hierarchical Bayesian approach is used to estimate an interval scaled measure of each household's utility for the eight items, and the household's price and promotion sensitivity. The retailer's decision problem is then solved as an integer programming problem. Although the problem is large in terms of the number of decision variables and constraints, we show that it can be solved efficiently. Our solution reveals that a significant reduction in customer disutility can be accomplished at the cost of a small reduction in the current period profits.

Our model should be considered as an illustrative first step. While we have captured the richness of customer heterogeneity, substitution behavior, and the current vs. future profit tradeoff, we also have made simplifying assumptions on other aspects of this complex problem. In Sect. 6 we outline several ways to enhance our proposed model to incorporate these remaining aspects, which we hope will inform further research in this important field.

The rest of the chapter is organized as follows. In Sect. 2 we review related research. We discuss the consumer model in Sect. 3. In Sect. 4 we develop an optimization framework for the assortment decision, discuss special cases and some properties of the model. In Sect. 5, we demonstrate an empirical application of our proposed model using household panel data. We conclude in Sect. 6 with a brief summary and a discussion of extensions and further research.

2 Literature Review

Two broad streams of literature are relevant to this study—one in marketing, the other in operations management. Early research in marketing deals with issues of retail shelf space allocation and is empirical in nature. Corstjens and Doyle (1981) proposed a model to optimize space allocation across categories, given an overall store space constraint. Direct and cross space elasticities were measured via a multiplicative sales response model using cross-sectional data. Their model does not explicitly include the assortment decision, although allocation of zero space to an item may be interpreted as exclusion of the item. However, as pointed out by Borin et al. (1994), the multiplicative sales response model predicts zero sales for a given category if the space of any of the store's other categories is set to zero. Bultez and Naert (1988) and Bultez et al. (1989) model the intra-category space allocation problem. Space elasticities are measured experimentally with item sales as the criterion variable. However, the assortment decision is not explicitly modeled. Borin et al. (1994) incorporate both the space allocation and assortment decisions in a retailer model. However, this study does not empirically estimate the demand model. Instead, parameter values are assumed. More recently, van Dijk et al. (2004) use observed variation in shelf-space allocation across stores to infer shelf-space elasticities.

The focus of these studies is on allocation of a scarce resource—space—given that different items show varying responsiveness to space. Thus, emphasis is placed on methods and data for measurement of space elasticities (own and cross) and on algorithms to solve the retailer profit maximization problem efficiently. By contrast, our focus is on estimating consumers' brand preferences to infer their willingness to substitute, thereby determining the optimal assortment of items to stock. In the present study we do not tackle issues of responsiveness of demand to space allocations, but leave that for future research. The primary emphasis in our work is motivated by the empirical observation that in most consumer packaged goods categories, consumers can often be (imperfectly) satisfied by one of several items. This characteristic of consumer behavior is used in determining optimal assortments.

Recent empirical findings in the marketing literature provide strong support for the idea that assortment reductions may be profitable for retailers. Broniarczyk et al. (1998) conduct controlled lab experiments as well as field experiments in which assortments were reduced in five categories in convenience stores. They measure consumer perceptions of variety, which are shown to mediate store choice.

A key finding is that elimination of low-selling items had little or no impact on shoppers' perceptions of variety, as long as favorite items were available and category shelf space was held constant.

Boatwright and Nunes (2001) analyze data from a natural experiment conducted by an online grocer, in which 94 % of the categories experienced dramatic reductions in the number of SKUs offered. Sales increased an average of 11 % across the 42 categories examined.[1] An important finding especially relevant to our work is that customers who lose their favorite item when the assortment is reduced are significantly less likely to purchase in the category on a future purchase occasion.

Borle et al. (2005) use household panel data of the same online grocer that Boatwright and Nunes study to analyze the effects of assortment reductions in several categories on overall store sales. They find that although the effect is positive in several categories, overall store sales are reduced due to decreases in the number of store visits and the size of the shopping basket. To our knowledge, this is the first study that demonstrates that customer retention, i.e., customers' repeat store visit behavior, is adversely affected by reductions in category assortments.

Sloot et al. (2006) distinguish between short and long term sales effects of a 25 % item reduction in the assortment in one category. They find that while short-term category sales suffer a sharp reduction, long-term category sales display only a weak negative effect.

The findings of both Broniarczyk et al. (1998) and Boatwright and Nunes (2001) highlight that the impact of assortment reductions is heterogeneous across consumers, depending on the extent of loyalty exhibited towards the lost item. Borle et al. (2005) show conclusively that assortment reductions may reduce a shopper's probability of returning to this store on the next shopping visit. Although our data do not permit us to directly model the effects of assortment availability on consumers' store choice decisions, in our assortment optimization model we formalize the idea by including in the retailer's objective function the disutility incurred by consumers as a result of not finding their preferred items in the available assortment. This disutility is idiosyncratic to each consumer, and serves as a proxy for the reduced profits resulting from the lower probability of consumers choosing this retailer in future.

In the operations literature, work on assortment problems was motivated by the textile industry where decisions regarding which sizes (e.g., in-seam lengths for slacks) to carry had to be made. Pentico (1974) considers the single dimension assortment problem with probabilistic demands, with assumptions about substitution behavior of consumers. Pentico (1988) extends the earlier work to two-dimensional assortment problems with deterministic demands. Other related work deals with determining optimal stock levels for multiple items given stochastic demands and a pattern of substitution based on non-availability; see, for example, Bassok et al. (1997)

[1] Part of the increase is attributed to enhanced utility due to reduced clutter in the category. Our model does not allow for such an effect.

and the references therein. In this work, however, substitution is determined by the supplier firm and not by the buyer or consumer.

van Ryzin and Mahajan (1999) study a stochastic single period assortment planning problem under a Multinomial Logit (MNL) Choice model. A consumer's choice depends on the variants that the store carries and they assume that consumers do not substitute in the event of a stock-out. Using a newsvendor framework with identical exogenous retail prices across all variants, they show that the optimal assortment always consists of a certain number of the most "popular" products. They also illustrate that retail prices and profits increase when consumer preferences are more "fashion" oriented. In a follow-up paper, Mahajan and van Ryzin (2001) incorporate both assortment-based as well as stock-out based substitution behavior and present a stochastic sample path optimization method to solve for the optimal assortment. In contrast to these papers that assume a MNL model of choice, Gaur and Honhon (2006) use a locational choice model to study the assortment problem.

Smith and Agrawal (2000) study the assortment planning problem using a general probabilistic model of demand allowing for substitution behavior. Using a substitution matrix, they estimate the derived demand for a given assortment. They then present a methodology to determine the assortment and stocking levels jointly when retailers incur a fixed cost for carrying an item in stock as well as the classical inventory and shortage costs for excess inventory and shortage at the end of the period.

Some recent papers have focused on jointly addressing demand estimation as well as assortment planning. Chong et al. (2001) present a category assortment planning problem. Consumer choice is represented as a combination of a category-purchase-incidence model and a brand-share model. While the former predicts the probability of an individual consumer's purchase from a category on a given shopping trip, the latter predicts which brand will be purchased. The optimization problem then determines the optimal number of facings for the various products to maximize profits, subject to a shelf space constraint. They illustrate their methodology using data from five stores in eight food categories.

Kok and Fisher (2004) present a demand estimation as well as an assortment optimization model. Using cross-sectional data across stores that carry different assortments, they estimate the substitution behavior of a homogenous set of customers. Using a probabilistic model of choice, they posit an assortment optimization model and develop heuristics to determine the number of facings of a particular product that a retailer should carry. They apply their method to a supermarket chain in the Netherlands and illustrate that their methodology for assortment planning potentially leads to a 50 % increase in profits.

Miller et al. (2006) propose an approach to optimize retail assortments with demand specified as a multinomial logit model. Consumers' utilities for products are estimated via a conjoint approach wherein consumer heterogeneity is allowed. In an empirical application they find that there is a significant negative impact on profits when heterogeneous consumers are assumed to be homogeneous.

Like the papers just discussed, our chapter focuses on a joint demand estimation and assortment planning problem. Demand is modeled at the household level using a discrete choice framework, specifically a probit model. Households are modeled as heterogeneous in unobserved utility function parameters, and the heterogeneity distribution is estimated using household scanner panel data. Thereby, posterior estimates of households' preference are derived.

The formulation of our optimization model is similar to the one studied by Dobson and Kalish (1988, 1993) in the context of positioning and pricing a product line. They present welfare and profit maximization formulations for positioning and pricing respectively. Our formulation is also similar to McBride and Zufryden (1988) who apply integer programming techniques to the optimal product line selection problem. Their model formulation recognizes heterogeneity in consumer preferences. Our approach of incorporating consumer disutility into the retailer's objective function is, however, more general than that of Dobson and Kalish (1993) or McBride and Zufryden (1988). The idea of penalizing the objective function for lost goodwill due to non-availability of stock is not new. In stochastic inventory theory (Arrow et al. 1958; Lee and Nahmias 1994) a penalty cost for shortages is routinely included in the objective function. However, to our knowledge, this chapter is the first to operationalize the penalty based on disutilities estimated from market-place data.

A key point of distinction between our paper and most of the literature discussed previously is with respect to the model of consumer heterogeneity. The classical multinomial logit (MNL) model as used in van Ryzin and Mahajan (1999) and Mahajan and van Ryzin (2001) allows for heterogeneity between consumers only via the stochastic term in the random utility. However, these differences between consumers are unobservable to the firm a priori, since the *expected* utility of a product is identical across consumers. This is why the model is sometimes referred to as the "homogeneous" MNL model. By contrast, we explicitly incorporate differences between consumers in the expected utility via a distributional assumption on the utility function parameters. The distribution of these parameters is then empirically estimated and can be used when determining the optimal assortment. Our approach is similar in theory to conjoint models (e.g., Miller et al. 2006) in which idiosyncratic utility functions are estimated.

3 Consumer Model

Our model of the retailer's decision problem of which items to carry and how much to carry, discussed at length in the next section, assumes that each consumer chooses that item from the available assortment which maximizes the consumer's utility. Solving this problem requires empirical estimates of consumers' preferences. We discuss in this section our approach to estimate consumer preferences.

Traditionally, data on consumer preferences have been collected via surveys as stated preferences (ordinal- or interval-scaled), or trade-offs that individuals would be willing to make on particular attributes (e.g., conjoint studies). An alternative approach is revealed preference data as obtained from reported or observed brand choices of consumers in actual purchase situations. For most product categories in the grocery industry these data are readily available from syndicated sources (e.g., household panels of Nielsen and Information Resources Inc.). The primary advantage of stated preference data is the ability to measure preferences for items currently not stocked (in particular, for new products). The major disadvantages of stated preference data relative to brand choice data are potentially lower validity of the data, and often substantially higher cost of data gathering.

Since the focus of our empirical work is on assortment decisions for supermarket product categories, we consider a model to estimate preferences that can be applied to observed brand choices of consumers—a multinomial probit model of brand choice. The probit model can be derived by assuming that the utility a consumer obtains from purchasing an item in the category is composed of a deterministic component and a stochastic component. The stochastic component represents unobserved (to the researcher) components of utility. In the typical formulation of the brand choice model, the utility of item $j, j = 1, 2, \ldots, J$ to consumer i on occasion t is given by U_{ijt}, thus: $U_{ijt} = \tilde{V}_{ijt} + \varepsilon_{ijt}, \varepsilon_{ijt} \sim N(0, \Sigma)$ where

$$\tilde{V}_{ijt} = \tilde{\alpha}_{ij} - \beta_i p_{ijt} + \tilde{\gamma}_i X_{ijt} \tag{10.1}$$

where for consumer i and item j, $\tilde{\alpha}_{ij}$ is the *intrinsic* utility or valuation, p_{ijt} is the price of the item on occasion t, X_{ijt} represents other attributes of the item (such as in-store promotions) on that occasion, and β_i and $\tilde{\gamma}_i$ are parameters. The assumption that the stochastic term has a multivariate normal distribution leads to the multinomial probit model of brand choice. We use a diagonal covariance structure $\varepsilon_{ijt} \sim N(0, \Sigma)$ where Σ is a $J \times J$ diagonal matrix, coupled with the identifying restriction that the first diagonal element is one. The choice of diagonal covariance structure simplifies the calculation of choice probabilities, while obviating the restrictive IIA property associated with a scalar covariance matrix, as well as with a multinomial logit model.

Note that the parameters of the utility function are individual specific, thus allowing for heterogeneity in both the intrinsic preferences and the effects of price and other attributes. As we demonstrate subsequently, this characteristic of the model has important implications for the optimal assortment decision of the retailer. The objective of model estimation is to recover the unknown parameters of the deterministic component of the utility function. Data required to estimate the model are observations of consumer choices as well as prices and in-store promotional conditions on each purchase occasion. Such information is typically available in household scanner panel data.

We model heterogeneity by specifying a series of conditional distributions in a Hierarchical Bayesian fashion. The reader is referred to Imai and van Dyck (2005) and McCullogh and Rossi (1994) for details of the estimation approach. A key benefit of using this approach is that it yields posterior estimates of utility function parameters at the individual level. These estimated utility functions are inputs into the retailer's optimization problem.

To obtain item-specific intrinsic utilities, we assume that prices are determined exogenously.[2] Furthermore, for simplification they are assumed to remain constant at their observed mean level p_j. We also assume the in-store promotion variables are fixed at their average levels X_j, again for simplification. Since utility is linear in prices, we divide utilities by the estimated price coefficient β_i (Kalish and Nelson 1991) to obtain a \$-metric utility, thus:

$$V_{ij} = \alpha_{ij} - p_j + \gamma_i X_j \tag{10.2}$$

where $\alpha_{ij} = \tilde{\alpha}_{ij}/\beta_i$, $\gamma_i = \tilde{\gamma}_i/\beta_i$ and p_j is the (constant) price of item j.[3]

The difference in \$-utility between two items may be considered the cost of substituting one item for the other for the consumer; see Krishna (1992) and Bawa and Shoemaker (1987) for a similar notion of substitution costs. An alternative interpretation of this difference is the reduction in price of the less preferred item necessary to make the consumer indifferent between the two items.

We assume that a consumer is willing to substitute lower utility items when higher utility items are not carried in the retail assortment. This assumption is strongly supported by empirical studies (Urban et al. 1984; Emmelhainz et al. 1991). The order of substitution is described by the rank-ordering of estimated preferences for items. When such substitution occurs, however, the consumer is assumed to incur a disutility equal to the difference in \$-metric of intrinsic utility between the most preferred item in the category and the item bought (i.e., the substitute item).

Empirical evidence also suggests that consumers may be willing to incur disutility due to downward substitution only upto a point. Below this point they may be unwilling to substitute and may choose to either postpone purchasing in the category or purchase at a different store (Borle et al. 2005). In an ideal setting, one would estimate the utility of a no-purchase decision and expect that consumers will be willing to substitute items as long as the utility of these items is above the utility for no-purchase. However, in the form they are currently available, household scanner panel data do not allow empirical estimation of the no-purchase threshold of households. Thus, in the subsequent empirical illustration we posit alternate mechanisms for operationalizing the no-purchase decision; we outline some options in Sect. 4.2.

Since the vector of intrinsic brand utilities is unique to each consumer, our consumer model allows completely idiosyncratic patterns of substitution. Not only is the highest preference brand allowed to be different across consumers, consumers who have a given brand as the most preferred may substitute a different brand in the event the most preferred item is not carried in the assortment. Such heterogeneity in substitution behavior between consumers has been documented in empirical

[2] In Sect. 6 of the chapter, as future research, we discuss the possibility of extending the model to determine optimal prices as well.

[3] The transformation of utilities by dividing by the price coefficient also serves to remove the influence of the unidentified scale factor that confounds the vector of parameter estimates (Swait and Louviere 1993).

studies (Emmelhainz et al. 1991). Furthermore, since we obtain an interval-scaled measure of preference, consumers who have exactly the same rank-ordering of brand preferences may incur differing amounts of disutilities due to non-availability of the most preferred item. This allows us to capture differences in intensities of brand preferences between consumers (e.g., loyals vs. switchers) that are relevant for the assortment and inventory decision.

To summarize, our model of the process consumers follow to choose an item to purchase in a category after entering the store is as follows. Consumers have preferences for various items in a category; these preferences vary from consumer to consumer. A consumer observes the available assortments (and the prices of items) and picks the highest utility item from those available or choses not to purchase. The exact operationalization of the no-purchase decision is discussed in the next section.

To use the consumer demand model in the retailer optimization problem, we revert to the utility measures V_{ij} in (10.2) (at constant prices) and use the estimated utilities $\widehat{V}_{ij}$. Disutilities form an important component of the retailer's objective function in our model, as detailed in the subsequent section. Ideally, we should use the random utility function U_{ijt} shown earlier. However, since U_{ijt} contains both a deterministic and a stochastic component, its use will lead to a potentially complex stochastic programming formulation. While accurate, this formulation does confound the impact of heterogeneity and probabilistic choice on the assortment decision. Instead, to focus exclusively on the heterogenous model of consumer behavior, we use only the deterministic component of the utility given by V_{ijt}. Our modeling choice is not without precedence; see Dobson and Kalish (1988, 1993) and McBride and Zufryden (1988). We comment on alternative approaches that could incorporate stochastic choice in the concluding section.

4 The Retailer Assortment and Stocking Problem

In this section, we describe a model to solve the retailer's assortment and stocking problem. We first develop a basic model that incorporates profits and disutility. We then discuss some special cases and properties of the formulation.

4.1 Basic Formulation

The retailer's problem can be defined as follows: We are given a set of N items indexed by j. There is a fixed cost of stocking each item. Consumers belong to one of the s index segments,[4] $s \in \{1, \cdots, S\}$. There exists a (monetary) utility

[4] The consumer model in Sect. 3 was developed assuming each consumer is a separate segment, i.e., the number of consumers in each segment is one. Other models of brand choice that provide estimates for "segments" of consumers could be employed, such as formulations of Kamakura and Russell (1989) and Chintagunta et al. (1991).

measurement, V_{sj}, for every segment s for every item j (see Sect. 3). As noted previously, for solving the retailer's optimization problem we assume that prices and promotional activities are held constant at their average levels. As a consequence, item utilities are time invariant. A consumer (segment) chooses from all available items the one that maximizes its utility.[5] The retailer's problem is to select an assortment and determine the stock for items in the assortment to maximize profits. The profit function can be written as:

$$PR(\mathbf{x}, \mathbf{y}) = \sum_j \left[\sum_s (p_j - c_j) x_{sj} n_s - K_j y_j \right] \tag{10.3}$$

where p_j is the per unit (regular) price of item j, c_j is the per unit variable cost of stocking item j, x_{sj} is a 0–1 variable which takes on a value of one if segment s customers are assigned to item j and zero otherwise (a decision variable),[6] n_s is the number of consumers in segment s, K_j is the fixed cost of stocking item j, and y_j is a 0–1 decision variable which takes the value one if item j is stocked and zero otherwise. Finally, $\mathbf{x}$ is a $S \times N + 1$ matrix of x_{sj} and $\mathbf{y}$ is an $N + 1$-vector of y_j. We let no-purchase decision be a "product" that is always available, thus expanding the product space to $N + 1$; further, $p_0 = c_0 = K_0 = 0$ and $y_0 = 1$.

Typically a retailer may do assortment planning for its stores twice a year; thus the planning horizon for assortments is about 6 months. In our formulation, we have not specified any planning horizon explicitly. The data can be scaled to accommodate any planning horizon. We need to, however, consider the fixed costs—which include costs relating to sourcing, supplier selection, negotiations, etc.,—appropriate for the planning horizon. Due to fixed costs of carrying an item in the assortment, not all items may be stocked. As a consequence, the following situations are possible:

1. A customer segment buys a less preferred item because its most preferred item is not available.
2. A customer segment does not purchase at all because no satisfactory item is available.

In either case the customer incurs a disutility. We postulate that such disutility adversely affects the customer's likelihood of repurchasing at this store, thereby affecting long-run profits.[7] We propose the following measure of customer disutility:

---

[5] We assume, for simplification, that each consumer buys exactly one unit in each restocking period. This assumption can be relaxed by weighting each consumer by the number of units bought. In general, the number of units bought by a consumer within any stocking period may be uncertain. Incorporating this uncertainty will result in a stochastic programming formulation. We elaborate upon this idea in the discussion of future work in Sect. 6.

[6] In the optimization model, the item "assigned" to a consumer will be the one that maximizes the consumer's utility. Thus, consumers will in effect self-select their best alternative from the available assortment.

[7] Notice that this disutility is due to non-stocking of items and not due to stock-out of an item.

$$DU(\mathbf{x}) = \sum_s n_s \left[\sum_k \{(V_{sj_1} - V_{sk})x_{sk}\} + (V_{sj_1} - V_{sj_0})(1 - \sum_k x_{sk}) \right] \tag{10.4}$$

where, $V_{sj_1} = \max_j\{V_{sj}\}$, and V_{sj_0} is the no-purchase utility, as discussed later in Sect. 4.2.

For those customers who are assigned an item k, the disutility is the difference between the utility of item k and their most preferred item.[8] Similarly, customers who do not purchase are also dissatisfied. The disutility incurred by these customers is the difference in utility between their highest utility and their utility for no-purchase. Clearly, customers who find their most preferred item in the assortment do not incur any disutility.

We propose that the overall objective function for a retailer should be a weighted combination of profits as measured by (10.3) and disutility as measured by (10.4). The extent to which a retailer should weight consumer disutility will depend on the product category. Customer dissatisfaction with some categories is likely to have a larger adverse impact on store choice. In the context of pricing, for example, Harris and McPartland (1993) classify categories into "traffic generators" (i.e., affect store choice) and others. We model this by taking a convex combination of the profit and disutility functions. Thus the objective function of the retailer is:

$$\Pi(\mathbf{x}, \mathbf{y}, w_c) = (1 - w_c)PR(\mathbf{x}, \mathbf{y}) - w_c DU(\mathbf{x}) \tag{10.5}$$

where $0 \leq w_c \leq 1$. w_c may be interpreted as a control or policy parameter whose value is to be subjectively determined by the decision maker.[9]

The optimization problem of the retailer is then written as follows:

$$\text{(P1)} \max_{\mathbf{x},\mathbf{y}} \Pi(\mathbf{x}, \mathbf{y}, w_c)$$

such that,

$$\sum_k V_{sk}x_{sk} \quad \geq \quad V_{sj}y_j \qquad \forall s, j \tag{10.6a}$$

[8] Dissatisfaction measured as sum across segments of the differences in utilities implies that a large number of small disutilities is equivalent to a small number of large disutilities; e.g., two segments with one unit of disutility each is equivalent to one segment (of same size) with two units of disutility. This may not be desirable since larger differences in utilities signify consumers *loyal* to certain brands, and smaller differences in utilities signify *switchers*. A non-linear (say, e.g., exponential) function of difference in utilities will allow us to distinguish between *loyals* and *switchers*.

[9] A similar objective function (weighted combination of profits and consumer utility) was also considered by Little and Shapiro (1980) in the context of pricing nonfeatured products in supermarkets. Similarly, there is extensive literature on bi-criterion optimization problems; see, for example, French and Ruiz-Diaz (1983).

$$\sum_j x_{sj} \leq 1 \qquad \forall s \tag{10.6b}$$

$$x_{sj} \leq y_j \qquad \forall s, j \tag{10.6c}$$

$$x_{sj} = 0, 1 \qquad \forall s, j \tag{10.6d}$$

$$y_j = 0, 1 \qquad \forall j \neq 0 \tag{10.6e}$$

$$y_0 = 1 \tag{10.6f}$$

Constraints (10.6a) ensure that of the items stocked, a customer is assigned his/her most preferred item. Constraints (10.6b) ensure that segment s is assigned to at most one item; finally, constraints (10.6c) ensure that only items that are offered are chosen by the customers.

At first glance it may appear that incorporating consumer disutility through $DU(\cdot)$ in the objective function makes constraints (10.6a) redundant. The constraints are redundant (or trivially satisfied) only when a retailer sets $w_c = 1.0$. Otherwise, in the absence of constraints (10.6a) it is possible that a retailer may assign a less preferred item (with a higher contribution margin) to a consumer even though a more preferred item (with a lower contribution margin) is stocked, albeit for a different consumer. Such an assignment is problematic from an implementation viewpoint in the context of supermarkets since a consumer walks into a store and necessarily picks his most preferred item if it is available. Constraints (10.6a) ensure that the retailer incorporates this fact into its decision making.

4.2 Modeling No Purchase

As discussed previously, a customer may decide to not purchase in the category if its preferred item is not stocked. Since scanner data do not report non-purchasing on account of unavailability in the assortment, we model this outcome and assume its value.[10] There are at least two ways one could model no purchase in the optimization problem. For a customer segment s, first rank order the utilities V_{sj} in decreasing order to write:

$$V_{sj_1} \geq V_{sj_2} \geq \cdots \geq V_{sj_N}.$$

[10] Category purchase incidence is frequently modeled using scanner data (e.g. Bucklin and Gupta 1992). However, the consumers' decision is considered to be one of choosing to buy one of the items in the assortment at today's prices and promotions, versus postponing the purchase decision to a future occasion when prices may be better, and relying meanwhile on available household inventory for consumption. Thus, the impact of assortment unavailability is not modeled.

Then,

1. For all customer segments s, assume that customers do not purchase if their most preferred d (exogenously specified) items are not stocked (see Smith and Agrawal 2000 for a similar operationalization). We call d the *depth of no purchase*. Clearly $d \in [1, N]$. An alternate interpretation of d is that it captures the (store) switching cost of a consumer; a large d implies high switching cost. Intuitively, a large d implies that a customer is willing to substitute less preferred items when more preferred items are not stocked rather than not purchase, regardless of the magnitude of disutility incurred. Under this operationalization, we set the no-purchase utility $V_{sj_0} = V_{sj_{d+1}}$ if $d < N$ and $V_{sj_0} = V_{sj_N} - \varepsilon$ (for some $\varepsilon > 0$) if $d = N$
2. Alternately, let T be an exogenously specified threshold level of disutility that signifies no purchase. Suppose there exists an item j_{k+1} for segment s, such that $V_{sj_1} - V_{sj_{k+1}} \geq T$. Then we infer that a customer in segment s will not purchase if items j_1 through j_k are not available in the assortment. Under this operationalization, we set the no-purchase utility $V_{sj_0} = V_{sj_{k+1}}$.

While either formulation is easily incorporated in our model, in this chapter, we use the former approach to model no-purchase. Later, we will analyze the sensitivity of the assortment solution to the depth of no purchase, d. To incorporate the depth of no purchase into problem P1, we modify constraint (10.6a) as follows. For each customer segment, s, define an order set consisting of d elements $\mathcal{N}_s^d = \{j_1, j_2, \ldots, j_d, j_0 | V_{sj_1} \geq V_{sj_2} \geq V_{sj_d} \geq V_{sj_0}\}$. We then rewrite (10.6a) as:

$$\sum_{k=0}^{d} V_{sj_k} x_{sj_k} \geq V_{sj_i} y_{j_i} \qquad \text{for } j_i \in \mathcal{N}_s^d \text{ and } \forall s \tag{10.6a$'$}$$

Furthermore, to ensure that a customer is assigned a product within their first d choices or no-purchase, we need to modify constraint (10.6c) to:

$$\sum_{j=0}^{d} x_{sj} \leq 1 \qquad \forall s \tag{10.6b1$'$}$$

$$\sum_{j=d+1}^{N} x_{sj} \leq 0 \qquad \forall s \tag{10.6b2$'$}$$

4.3 Reformulation

In this section we reformulate problem (P1), specifically constraint (10.6a$'$) which facilitates solution of (P1) as a linear program when integrality constraints on x_{sj} are relaxed. We observe that constraint set (10.6b)–(10.6e) is of the same form as that for an uncapacitated plant/warehouse location problem (Cornuejols et al. 1977). We now reformulate constraint set (10.6a′) that results in a tighter formulation for (P1). Observe that (10.6a′) ensures that a customer segment is assigned its most preferred product amongst the ones stocked. Thus it merely depends on the rank order of products for any given consumer segment and not on the interval scaled utilities as measured by V_{sj}. We exploit this structure to replace (10.6a′) with.

$$1 - \sum_{k=i+1}^{d} x_{sj_k} \geq y_{j_i} \qquad \text{for } j_i \in \mathcal{N}_s^d \text{ and } \forall s \tag{10.6a″}$$

We also relax the constraints on x_{sj} in (10.6d) as follows:.

$$x_{sj} \leq 1 \tag{10.6d′}$$

Proposition 4.1. *Problem (P1) with (10.6a″) set of constraints is at least as tight a formulation as (P1) with (10.6a′) set of constraints. Furthermore the relaxation of integrality constraints to (10.6a′) still guarantees an integer solution for* x_{sj}.

A proof is provided in the appendix.

Thus the new constraint set (10.6a″) achieves the same results as (10.6a′), i.e., ensuring that of the items stocked a customer segment is assigned its most preferred item. Furthermore, this reformulation does not increase the number of constraints. Finally, the relaxation guarantees an integer solution. In the sequel we will use (P1) with (10.6a″) and (10.6d′).

4.4 Discussion of the Optimization Model and Some Special Cases

Readers familiar with the literature on plant location will see that problem (P1) has an embedded uncapacitated plant location model (when $w_c = 0$, and constraints (10.6a) are relaxed). This problem is extensively researched by Cornuejols et al. (1977) and they show that the problem is NP-hard. Hence problem (P1) is also NP-hard. Our computational study shows that similar to the uncapacitated plant location model (Erlenkotter 1978), the solution to problem (P1) is easily obtained for problem sizes (relatively small) of interest in this study. Large scale models comprising several products in a product line and a larger number of customer segments will call for development of heuristics.

We now consider a few special cases of Problem P1. First, we consider the situation when a retailer places zero weight on the disutility incurred by the consumers due to his assortment decision; we shall identify a retailer with $w_c = 0.0$ as a *myopic retailer* who maximizes just short-term profits.

To highlight the need to model “no purchase”, consider the myopic retailer who solves P1 with $w_c = 0.0$ and with a depth of no purchase $d < N$. Recall that as d increases, consumers are more willing to substitute to the available items in the assortment and less willing to not purchase. We then observe that in a model without a no-purchase decision, a myopic retailer will stock only one product. Effectively, we solve problem P1 with $w_c = 0.0$ and $d = N$; that is, the retailer does not care about disutilities incurred by the consumers and all consumers purchase some product. This implies that the total demand is unaffected by the choice of items available. Then a retailer carries just one product $j^* = argmax_j\{(p_j - c_j)n_s - K_j\}$ which maximizes his profit.

We would like to be able to study the behavior of the assortment decision with respect to parameters like weight on disutility (w_c), depth of no-purchase (d), contribution margins ($p_j - c_j$), etc. In general, (P1) is a complex optimization problem and usually does not permit many comparative statics results. Analytically, we were unable to get any general sensitivity results with respect to p_j, w_c and d. The main difficulty appears to be the very general formulation of the heterogeneity of consumers. Any change in these parameters affects the substitution pattern through change in the interval scaled utilities and hence the demand patterns. The obvious case is when profit margins increase due to decrease in marginal costs. This increases the contribution margin and with fixed p_j, d and w_c, the retailer will find it optimal to increase his assortment sizes, since for $d < N$ it may help him satisfy more consumers and/or decrease disutility if $w_c > 0$.

5 Computational Study

5.1 Description of Household Scanner Panel Data

The data were collected by the AC Nielsen Company and are available for a 2 year period. A panel of households provided information on their purchasing in several categories. These data were supplemented with data on prices, in-store displays, and feature advertising collected from the supermarkets in the city. We include purchases of the eight largest brand-sizes of canned tuna made by 1,097 panelist households in our estimation sample. These eight items account for approximately 90 % of category volume. Brand names are disguised to meet confidentiality requirements of the data provider.

In Table 10.1 we provide descriptive statistics of the data. Besides shelf price, we include in-store displays and retailer feature advertising in the choice model. Table 10.1 indicates that there is considerable variation in shelf prices and promotional activity between brands, highlighting the need to control for the effects of these variables when measuring intrinsic brand preference or valuation.

Bayesian posterior estimates of the model parameters are obtained for each household using the approaches of Imai and van Dyck (2005) and McCullogh and Rossi (1994). Table 10.2 contains the mean value of the estimated posterior estimates. The coefficients of price, display, and feature, have the expected signs.

Table 10.1 Descriptive statistics of data

Item	Average price (cents/oz.)	Display (% occasions)	Feature (% occasions)
1	12.3	3.9	25.9
2	21.8	0	1.7
3	12.0	4.0	29.9
4	11.5	8.7	24.4
5	15.1	0	0
6	24.2	0	0
7	11.3	4.3	24.4
8	9.8	4.2	13.7

Table 10.2 Mean value of household parameter estimates of probit model demand

Mean brand specific constants	
Item 1	0.815
Item 2	2.350
Item 3	0.494
Item 4	1.030
Item 5	0.267
Item 6	2.885
Item 7	−0.273
Price ($/oz.)	−26.882
Display	0.597
Feature	0.163

We use the estimated $-metric intrinsic preferences for items V_{ij} to infer patterns of primary demand and likely substitution between items. We computed optimal assortments under two separate assumptions about consumers' willingness to substitute. First we assume that consumers are willing to make one substitution. That is, they will not purchase in the category if their first preference and second preference brands are not available (i.e., $d=2$). Therefore, we focus on the top two brands for each consumer. Note that customers who do not find their most preferred brand but do find their second-most-preferred brand still incur a disutility, which our decision model incorporates. Next, we also solved for the optimal assortment under the assumption that consumers are willing to substitute twice (i.e., $d=3$). In the subsequent discussion we describe the solution under the $d=2$ assumption in detail and thereafter briefly talk about the $d=3$ case.

Table 10.3 shows the cross classification of the first and second preference brands for the sample of 1,097 consumers.[11] Row total $N_{i.}$ indicates the number

[11] Note that only the rank ordering of preferences is used to construct Table 10.3 to illustrate the nature of substitution between items. The retailer optimization problem uses interval-scaled values of preferences.

Table 10.3 Cross classification of first and second preference brands (cell entries are in %)

First preference product	Second preference product								No. of consumers
	1	2	3	4	5	6	7	8	
1	0.00	0	71.7	3.3	1.7	0	10.0	13.3	60
2	0	0	0	0	0	0	0	0	0
3	69.7	1.4	0	1.9	0	0	17.1	9.9	211
4	3.4	0	6.8	0	1.1	0	78.4	10.2	88
5	12.5	6.3	3.1	0	0	0	28.1	50.0	32
6	0	0	0	0	0	0	0	0	0
7	2.3	0	14.7	55.4	2.5	0	0	25.1	354
8	14.5	0	21.3	2.8	13.6	0	47.7	0	352
Number of consumers	213	5	177	212	59	0	288	143	1,097

of consumers whose first preference brand is brand i. Similarly, column total $N_{.j}$ is the number of consumers whose second preference brand is brand j. Each cell entry in the table denotes the percentage of $N_{i.}$ consumers who have brand j as their second preference brand.

The row totals are indicative of primary demands for items. For example, it is clear that items 3, 7 and 8 are the first-preference products of a large number of consumers, while none of the consumers in our sample prefer items 2 and 6. Similarly, items 1, 4, and 5 have relatively weak primary demand. Column totals indicate whether items are acceptable as substitutes. Item 1, for example, is the brand of second choice for a large number of consumers (213) as compared with its primary demand (60). A similar preference pattern is evident for items 3 and 4. Item 8 has the opposite kind of preference pattern, with large number of consumers (352) preferring it in first place while only 143 prefer it in second place. Large cell entries indicate items that are more substitutable. For example, we see that 71.7 % of consumers who have item 1 as their first preference have item 3 as their second preference. Conversely, 69.7 % of those who prefer item 3 are willing to accept item 1. There is some evidence of asymmetries in patterns of substitution between brands. For instance, the entry in row 5 and column 8 is 50.0 % while that in row 8 and column 5 is only 13.6 %. These data further confirm the existence of substantial heterogeneity in patterns of substitution between consumers.

5.2 Solution Technique for Assortment Problem

We used LINDO, a commercial linear programming package, to solve the reformulated optimization model. The problems are generated from the preference, price, and cost data using a program written in C. This program allows the decision maker to vary the weight w_c (weight on consumer welfare and profit objectives) and d (depth of no purchase) to evaluate various solutions.

For our computational study we solved 80 instances of the problem. We varied the weight w_c from 0.01 to 0.99 with $d = 2$ (40 problems) and d = 3 (40 problems) for two different fixed costs. On average the problem took 32 s of cpu time, with times ranging from 20 to 48 s. Based on our computational times it seems appropriate to solve this problem to obtain the optimal solution using a commercial package. Specialized implementation and heuristics may be necessary for larger problems if the computational times become prohibitive.

5.3 Optimal Assortment

To solve the retailer optimization problem (P1), we need estimates of fixed costs (K_j), contribution margins ($p_j - c_j$), and of w_c, the weight placed by the retailer on customer disutility relative to current period profits. We did not have access to real cost and contribution data for the market for which consumer data were available. For the empirical illustration, we assume values of these parameters as follows. Retail contribution margins are assumed to be 30 % of the average retail price of the item. Thus, items can be ordered in terms of margin based on the average prices shown in Table 10.1. We examine two different levels of fixed costs in our illustrations: \$1 per re-stocking period and \$5 per stocking period. These levels of fixed costs ensure that at least one item is unprofitable to carry based on its primary demand. We explore the impact of varying w_c (over the space 0 to 0.99 in small steps) on the optimal assortment, profits and customer disutility.[12]

Case 1: Fixed Cost is \$1 per item per stocking period

In Table 10.4 we show changes in the optimal assortment of items, customer disutility, and optimal profits as the weight on disutility in the objective function (w_c) is increased from 0 to 0.99. Note that items 2 and 6 are never included in the optimal assortment, regardless of the value of w_c, because of the pattern of first and second preferences discussed previously. When $w_c = 0$, the problem reduces to the pure profit maximization problem of a myopic retailer. Thus, the retailer should carry only those products whose contribution margin exceeds the fixed cost. The demand for a product, given an assortment, is the sum of its primary demand, and spillover demand from items not carried. The solution to the pure profit maximization problem is to carry four items (item numbers 1, 3, 5, and 7). Table 10.1 shows that products 1, 3, and 5 are the highest margin products (after products 6 and 2). Although item 4 has higher margin than item 7, item 7 is included in the optimal assortment instead of item 4 because of its large primary demand (354 consumers) relative to item 4 (88 consumers). When a weight of 0.03 is placed on disutility we find that item 4 is also included in the assortment now. As noted previously, item 4 has low primary demand, but is acceptable as a

Table 10.4 Optimal assortment and resulting disutility and profits (fixed cost = \$1)

Weight on disutility (w_c)	Disutility	Profit	# customers not served	Optimal assortment
0.000	85.83	34.50	19	1, 3, 5, 7
0.010	85.83	34.50	19	1, 3, 5, 7
0.030	69.62	34.17	0	1, 3, 4, 5, 7
0.040	20.38	32.64	13	3, 7, 8
0.050	4.45	31.93	7	3, 4, 7, 8
0.200	2.27	31.60	0	3, 4, 5, 7, 8
0.300–0.990	0.00	30.72	0	1, 3, 4, 5, 7, 8

[12] For the illustration here we assume that the total market consists of the 1,097 consumers in our sample.

substitute by a large number of customers. Nineteen customers who were previously not served at all now find an acceptable product to buy. Moreover, with this assortment profits are slightly lower, but disutility is significantly reduced. This suggests that profit as a function of assortment carried is quite flat near the maximum. The introduction of a second criterion (i.e., disutility) into the objective function helps us to select the assortment that delivers close to maximum profits while reducing disutility. If customer disutility influences future store traffic and hence long-run profits, the results presented help the decision maker balance short-run with long-run profits.

As w_c is increased further, we find that the number of items in the optimal assortment decreases and then increases. At $w_c = 0.040$ the optimal assortment shrinks from {1,3,4,5,7} to {3,7,8}. The inclusion of item 8 is probably explained by its large primary demand (352 customers), which implies that when it is omitted from the assortment, large disutility is incurred. Further, half of the customers who prefer item 5 find item 8 acceptable. At $w_c = 0.050$ the optimal assortment expands to include item 4 once again. At $w_c = 0.30$ the optimal assortment expands to include all six products, other than items 2 and 6.

Note that we observe two kinds of non-monotonicities in the optimal behavior with increases in w_c. One, the number of items in the optimal assortment expands and then shrinks. Two, certain items (such as 4 and 1) enter the optimal assortment, then get dropped, and then get re-included. Such non-monotonic behavior of the optimal assortment reinforces the need for a decision support model for retail assortment decisions.

Case 2: Fixed Cost is $5 per item per stocking period

In Table 10.5 we show the optimal assortment and associated profits and disutility. Note that in the pure profit maximization case, 155 customers are not served and disutility incurred is quite high. Placing a weight of 0.03 on disutility expands the optimal assortment to include product 8 in addition to items 3 and 7. As a consequence, profits drop. However, the number of customers served increases significantly and disutility drops sharply.

A distinguishing feature of the optimal assortment in Case 2, relative to Case 1, is that with increase in w_c the number of items in the optimal assortment always increases. Furthermore, once an item enters the optimal assortment it stays in the

Table 10.5 Optimal assortment and resulting disutility and profits (fixed cost = $5)

Weight on disutility (w_c)	Disutility	Profit	# customers not served	Optimal assortment
0.000	95.92	22.72	155	3, 7
0.010	95.92	22.72	155	3, 7
0.030	20.38	20.64	13	3, 7, 8
0.300	4.45	15.93	7	3, 4, 7, 8
0.700–0.990	0.00	6.72	0	1, 3, 4, 5, 7, 8

Table 10.6 A three-product example

Customer	Utility		
	Product 1	Product 2	Product 3
1	5	2	1
2	5	2	1
3	1	2	5
Fixed cost	2	2	2
Margin	1.4	1.4	1.4

assortment with increases in w_c. We conjecture that the high fixed cost may cause such monotonic behavior of the optimal assortment.

Results in the $d = 3$ case are entirely consistent with the results for the $d = 2$ case with some differences that are intuitive. For reasons of space we do not show detailed results. At each level of w_c, we find that optimal profits are at least as large in the $d = 3$ case since consumers are assumed to be more willing to substitute to less-preferred products. As a result, the spillover demand to any product from items not carried is no lower in this case than in the $d = 2$ case. Further, disutility is at least as large in the $d = 3$ case. When the fixed cost per item is \$1, the optimal assortment changes non-monotonically with increases in w_c. When the fixed cost is \$5, on the other hand, the optimal assortment changes monotonically.

To deduce further inferences, we ran the model for both cases of fixed costs considered previously ($K = 1$ and 5) and equal margins across all products, set equal to average margin of eight products using depths $d = 2$ and $d = 3$. The optimal solutions exhibited monotone changes to the optimal assortment for all w_c values. While this is true for our particular data set, we are able to construct a three-product, three-customer instance to provide a counter-example (see data in Table 10.6) for this monotone behavior.

In this counterexample, we find that when $w_c = 0$, the optimal profits are 2.2, the disutility is 9, and the optimal assortment has only product 2. As w_c grows to 0. 1379, the assortment consists of product 1 only, and for higher values of w_c the optimal assortment consists of products 1 and 3.

The results show that it is very hard to predict the structure of the optimal assortment, especially when we consider a data-driven problem setting.

6 Summary, Extensions, and Future Work

We propose a model for the optimal assortment and stocking decisions for retail category management. In particular, we address the question of rationalization of the retail assortment, i.e., determining the optimal subset of items to retain from the set of items currently carried. We assume, based on empirical evidence reported in the literature, that consumers are willing to partially substitute less preferred items if their preferred items are not available. We also assume that consumers are heterogeneous in their intrinsic preferences for items and in their price sensitivities, an assumption strongly supported empirically.

We propose that the appropriate objective function for a far-sighted retailer should include not only short-term profits but also a penalty for the disutility incurred by consumers who do not find their preferred items in the available assortment. The rationale for including such a penalty is that dissatisfied consumers are less likely to return to the store in the future. We propose a measure for disutility that recognizes differences between consumers in their intensity of dissatisfaction.

The retailer problem is formulated as an integer programming problem. We show that the problem is large but can be solved efficiently to obtain an optimal solution. We demonstrate an empirical application of our proposed model using household scanner panel data for eight items in the canned tuna category. Our results indicate that the inclusion of the penalty for disutility in the retailer's objective function is informative in terms of choosing an assortment to carry. We find that customer disutility can be significantly reduced at the cost of a small reduction in short term profits.

An immediate extension of the current work is to develop heuristics to solve the optimization problem since problem sizes in categories with a large number of items may be very large and computational times to find optimal solutions might be prohibitive. Furthermore, we realize that there is uncertainty due to errors in the utility function parameter estimates, which our optimization model assumes to be fixed. The problem formulation can be modified to allow for uncertain parameter estimates and use a stochastic programming approach to solve the assortment problem.

The approach described in this chapter is an illustrative first-step that attempts to close some of the modeling gaps in the literature. As outlined in the introduction, the complete assortment planning problem needs to consider several other factors. Next we discuss briefly several directions to extend the proposed model in future research.

1. *Shelf Space Constraints:* Typically, retailers have shelf space constraints which limit the amount of stock that can be carried within a category. These constraints can be incorporated within the context of our problem (P1). A complexity that now arises is the occurrence of stock-outs. Since customers have heterogeneous preferences for items, the dynamics of their arrival process also needs to be accounted for.
2. *Incorporating Demand Uncertainty*: In the current model, we assumed that utilities of each consumer segment are deterministic. In fact, from the retailer's perspective utilities are stochastic. Including stochastic utilities results in a mixed-integer stochastic programming problem.
3. *The Pricing Problem*: The basic formulation outlined in this chapter can be extended to study the joint pricing and assortment decisions. However, maximization over prices makes (P1) a non-linear optimization problem which can be solved using procedures outlined in Adams and Sherali (1990), for example. Alternately, heuristic procedures could be explored.
4. *The Display Effect or the Effect of Facings on Sales*: The literature on shelf space management has been concerned with the relationship between shelf space allocations and sales due to the influence of product display on demand.

The number of facings allocated to an item also determines the quantity stocked of this item (usually an integer multiple of the number of facings). Thus, the problem of determining the optimal assortment and inventory is inter-related with the shelf-space allocation problem. Extending the model presented in this chapter to incorporate the display effect presents two challenges: one, the problem of measuring the effect of product display on demand, and two, the optimization problem changes considerably since we will now have to decide on number of facings which will be an integer variable.

5. *Joint Fixed Costs*: Product lines for a retailer typically consist of several SKU's being supplied by the same manufacturer or wholesaler. Therefore, multiple products in a category may require common resources (contact, vendor management, etc.). The Dobson and Kalish (1993) formulation assumes independent fixed costs, and therefore it can overstate the fixed costs associated with incremental introduction of products that share fixed costs with incumbent products. In case of shared fixed costs, a firm can take the savings available into account when introducing products that require common resources. One approach is to define product classes, similar to manufacturing classes used by Morgan et al. (2001). We hypothesize that inclusion of common fixed costs (relative to the assumption of independent fixed costs) will increase the number of products offered, profits, as well as consumer satisfaction.

Acknowledgements We are grateful to the A.C. Nielsen Company for generously providing the data used in this paper, to Edward Malthouse for his help with setting up the data, to Qiang Liu for help with data analysis, and to Pradeep Chintagunta, Maqbool Dada and Yehuda Bassok for valuable comments on an earlier version of the paper.

Appendix

Proof of Proposition 4.1. Without loss of generality, we will illustrate this for the general case rather than the special case of fixed depth of search d.

First consider the constraint (10. 6a$''$). The constraint for $j = 1$ will be

$$1 - (x_{s2} + x_{s3} + \ldots + x_{sK} + x_{s0}) \geq y_1$$

However, from (10.6c) we know that

$$x_{s1} + x_{s2} + \ldots + x_{jK} + x_{s0} \leq 1$$

Actually, given a "no purchase" option, the above is an equality; i.e.,

$$x_{s1} + x_{s2} + \ldots + x_{jK} + x_{s0} = 1$$

Using this we rewrite $1 - (x_{s2} + x_{s3} + \ldots + x_{sK} + x_{s0}) \geq y_1$ as simply $x_1 \geq y_1$. Similarly, we can write (10. 6a″) for $j = k$ as

$$x_{s1} + x_{s2} + \ldots + x_{sk} \geq y_k.$$

Using this, for any arbitrary customer segment s that prefers K products in the ordinal order (without loss of generality) the constraint sets (10. 6a′) and (10. 6a″) are

(10. 6a′)		(10. 6a″)	
$V_{s1}x_{s1} + V_{s2}x_{s2} + \cdots + V_{sk}x_{sk} \geq V_{s1}y_1$	$(1')$	$x_{s1} \geq y_1$	$(1'')$
$V_{s1}x_{s1} + V_{s2}x_{s2} + \cdots + V_{sk}x_{sk} \geq V_{s2}y_2$	$(2')$	$x_{s1} + x_{s2} \geq y_2$	$(2'')$
$\vdots$		$\vdots$	
$V_{s1}x_{s1} + V_{s2}x_{s2} + \cdots + V_{sk}x_{sk} \geq V_{sk-1}y_{k-1}$	$((k-1)')$	$x_{s1} + x_{s2} + \cdots + x_{sk-1} \geq y_{k-1}$	$((k-1)'')$
$V_{s1}x_{s1} + V_{s2}x_{s2} + \cdots + V_{sk}x_{sk} \geq V_{sk}y_k$	(k')	$x_{s1} + x_{s2} + \cdots + x_{sk} \geq y_k$	(k'')

Consider normalized constraint (1′) and (1″):

$$x_{s1} + \left(\frac{V_{s2}}{V_{s1}}\right)x_{s2} + \cdots \left(\frac{V_{sk}}{V_{s1}}\right)x_{sk} \qquad \text{and} \qquad x_{s1} \geq y_1.$$

Since $(1'')$ and $(1')$ are identical in x_{s1} dimension and $(1')$ has $k-1$ extra variables (degrees of freedom), constraint $(1'')$ is tighter than constraint $(1')$. Using similar arguments one can show that constraints $(2'')$ to $((k-1)'')$ will be tighter than $(2')$ to $((k-1)')$. Constraint (k'') may be identical to (k'). The argument can be repeated for other segments. Thus problem (P1) with (10. 6a″) is a tighter formulation than (P1) with (10. 6a′).

To see that relaxation of x_{sj} still leads to an integer solution, first consider $(1'')$. If $y_1 = 0$, then $x_{s1} = 0$ using (10.6c). If $y_1 = 1$, then $x_{s1} = 1$. Now consider $(2'')$. Suppose $y_1 = 0$. If $y_2 = 0$ then $x_{s2} = 0$; otherwise $(y_2 = 1)$, $x_{s2} = 1$. However, if $y_1 = 1$, then (10.6b) ensures that $x_{s2} = 0$. Following this argument, we can show that x_{sj} is integer. ■

References

Adams, P. W., & Sherali, H. D. (1990). Linearization strategies for a class of zero–one mixed integer programming problems. *Operations Research, 38*, 217–226.

Arrow, K., Karlin, S., & Scarf, H. (1958). *Studies in the mathematical theory of inventory and production*. Stanford: Stanford University Press.

Bassok, Y., Anupindi, R., & Akella, R. (1997). Single period multi–product inventory models with substitution. *Operations Research, 47*, 632–642.

Bawa, K., & Shoemaker, R. W. (1987). The effects of a direct mail coupon on brand choice behavior. *Journal of Marketing Research, 24*, 370–376.

Boatwright, P., & Nunes, J. C. (2001). Reducing assortment: An attribute based approach. *Journal of Marketing, 65*, 50–63.

Borin, N., Farris, P., & Freeland, J. (1994). A model for determining retail product category assortment and shelf space allocation. *Decision Sciences, 25*(3), 359–384.

Borle, S., Boatwright, P., Kadane, J. B., Nunes, J. C., & Shmueli, G. (2005). Effect of product assortment changes on customer retention. *Marketing Science, 24*(4), 612–622.

Broniarczyk, S. M., Hoyer, W. D., & McAlister, L. (1998). Consumers' perceptions of the assortment offered in a grocery category: The impact of item reduction. *Journal of Marketing Research, 35*, 166–176.

Bucklin, R. E., & Gupta, S. (1992). Brand choice, purchase incidence, and segmentation: An integrated modeling approach. *Journal of Marketing Research, 29*(9), 201–215.

Bultez, A., Gijsbrechts, E., Naert, P., & Vanden Abeele, P. (1989). Asymmetric cannibalism in retail assortments. *Journal of Retailing, 65*(2), 153–192.

Bultez, A., & Naert, P. (1988). S.H.A.R.P.: Shelf allocation for retailer's profit. *Marketing Science, 7*(3), 211–231.

Carpenter, G., & Lehmann, D. R. (1985). A model of marketing mix, brand switching and competition. *Journal of Marketing Research, 22*, 318–329.

Chintagunta, P., Jain, D. C., & Vilcassim, N. J. (1991). Investigating heterogeneity in brand preferences in logit models for panel data. *Journal of Marketing Research, 28*, 417–428.

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

Cornuejols, G., Fisher, M., & Nemhauser, G. (1977). Location of bank accounts to optimize float: An analytic study of exact and approximate algorithms. *Management Science, 23*(8), 789–810.

Corstjens, M., & Doyle, P. (1981). A model for optimizing retail space allocations. *Management Science, 27*(7), 822–833.

Coughlan, A. T., Anderson, E., Stern, L. W., & El-Ansary, A. I. (2006). *Marketing channels*. Englewood Cliffs, NJ: Prentice Hall.

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

Dobson, G., & Kalish, S. (1993). Heuristics for pricing and positioning a product–line using conjoint and cost data. *Management Science, 39*, 160–175.

Emmelhainz, M. A., Stock, J. R., & Emmelhainz, L. W. (1991). Consumer responses to retail stockouts. *Journal of Retailing, 67*(2), 139–147.

Erlenkotter, D. (1978). A dual-based procedure for uncapacitated facility location. *Operations Research, 26*(6), 992–1009.

Food Marketing Institute. (1993). *Variety or duplication: A process to know where you stand*. Washington, D.C.: The Research Department, Food Marketing Institute.

French, S., & Ruiz-Diaz, F. (1983). A Survey of multi–objective combinatorial scheduling. In S. French, et al. (Eds.), *Multi–objective decision making*. New York: Academic.

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

Gruen, T., Cortsen, D. S., & Bharadwaj, S. (2002). *Retail out-of-stocks: A worldwide examination of extent, causes and consumer responses*. Grocery Manufacturers of America.

Harris, B., & McPartland, M. (1993). Category management defined: What it is and why it works. *Progressive Grocer, 72*(9), 5–8.

Imai, K., & van Dyck, D. A. (2005). A Bayesian analysis of the multinomial probit model using the marginal data augmentation. *Journal of Econometrics, 124*(2), 311–334.

Kalish, S., & Nelson, P. (1991). A comparision of ranking, rating, and reservation price measurement in conjoint analysis. *Marketing Letters, 2*(4), 327–335.

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

Kok, A., & Fisher, M. L. (2004). Demand estimation and assortment optimization under substitution: Methodology and application. *Operations Research, 55*(6), 1001–1021.

Krishna, A. (1992). The normative impact of consumer price expectations for multiple brands on consumer purchase behavior. *Marketing Science, 11*(3), 266–286.

Lee, H. L., & Nahmias, S. (1994). Single product single location models. In S. Graves, A. R. Kan, & P. Zipkin (Eds.), *Logistics of production and inventory. Handbook in operations research and management science*. Amsterdam: North–Holland.

Little, J. D., & Shapiro, J. (1980). A theory for pricing nonfeatured products in supermarkets. *Journal of Business, 53*(3), S199–S209.

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

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

McCullogh, R., & Rossi, P. E. (1994). An exact likelihood analysis of the multinomial probit model. *Journal of Econometrics, 64*, 207–240.

Miller, C., Smith, S. A., McIntyre, S. H., & Achabal, D. D. (2006). *Optimizing retail assortments for infrequently purchased products*. Working Paper, Retail Management Institute, Santa Clara University.

Morgan, L. O., Daniels, R. L., & Kouvelis, P. (2001). Marketing/manufacturing trade-offs in product line management. *IIE Transactions, 33*, 949–962.

Nielsen Marketing Research (1992). *Category management: Positioning your organization to win*. Lincolnwood, IL: NTC Business Books.

Pentico, D. (1974). The assortment problem with probabilistic demands. *Management Science, 21*, 286–290.

Pentico, D. (1988). A discrete two-dimensional assortment problem. *Operations Research, 36*(2), 324–332.

Schiller, Z., Burns, G., & Miller K. L., (1996). Marketing: Making it simple. *Business Week*, 9 September.

Sloot, L., Fok, D., & Verhoff, P. C. (2006). The short- and long-term impact of an assortment reduction on category sales. *Journal of Marketing Research, XLIII*, 536–548.

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

Swait, J., & Louviere, J. (1993). The role of the scale factor in the estimation and comparision of multinomial logit models. *Journal of Marketing Research, 30*, 305–314.

Urban, G. L., Johnson, P. L., & Hauser, J. R. (1984). Testing competitive market structures. *Marketing Science, 3*(2), 83–112.

van Dijk, A., van Heerde, H. J., Leeflang, P. S. H., & Wittink, D. R. (2004). Similarity based spatial methods to estimate shelf-space elasticities. *Quantitative Marketing and Economics, 2*, 257–277.

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