Marketing Letters 16:3/4, 239–254, 2005

Models of Multi-Category Choice Behavior

P. B. SEETHARAMAN * seethu@rice.edu
Campus Box 2932, MS 531, Rice University, Houston, TX 77252

SIDDHARTHA CHIB
Washington University in St. Louis

ANDREW AINSLIE
University of California at Los Angeles

PETER BOATWRIGHT
Carnegie Mellon University

TAT CHAN
Washington University in St. Louis

SACHIN GUPTA
Cornell University

NITIN MEHTA
University of Toronto

VITHALA RAO
Cornell University

ANDREI STRIJNEV
University of Texas at Dallas

Abstract

Given the advent of basket-level purchasing data of households, choice modelers are actively engaged in the development of statistical and econometric models of multi-category choice behavior of households. This paper reviews current developments in this area of research, discussing the modeling methodologies that have been used, the empirical findings that have emerged so far, and directions for future research. We also motivate the use of Bayesian methods to overcome the computational challenges involved in estimation.

Keywords: multi-category, multivariate choices, basket data, bayesian estimation

1. Introduction

Multi-category models are of interest to not only retailers seeking to maximize store profits by jointly coordinating marketing activities across product categories, but also packaged

* Corresponding author

goods manufacturers such as Procter and Gamble, who sell products in multiple product categories (such as detergents, shampoos etc.). Such models are also of interest to firms, such as financial services providers, interested in undertaking cross-selling initiatives across product categories (Kamakura et al., 1991). Recent availability of *basket data* on consumer purchasing behavior in supermarkets from syndicated data providers such as IRI and A.C. Nielsen has facilitated the estimation of such models. For example, one such dataset—introduced to the marketing literature by David Bell at Wharton—tracks purchases of 494 households in 25 product categories at a set of 5 stores in a local market over a period of two years from June 1991 to June 1993. Such a dataset yields longitudinal (across shopping trips) information, at the household level, on four outcome variables within each product category: 1. store visited ("store choice"), 2. whether or not the product category was purchased ("incidence"), 3. which brand (if any) of the product was purchased ("brand choice"), and 4. how many units (if any) of the product were purchased ("quantity").

Choice modelers have devoted a good part of the past four decades in developing (statistical or econometric) models of these four purchase outcomes, in all possible subsets (taken one, two, or more outcomes at a time etc.). However, the majority of these models focus on *one product category at a time*. For example, choice models have been developed to model three outcome variables—incidence, brand choice and quantity—simultaneously *within a single product category* (see Chiang, 1991). In contrast, this paper focuses on modeling multiple purchase outcome variables simultaneously *across multiple product categories* (for earlier reviews, see Russell et al., 1997, 1999; Elrod et al., 2002). In other words, our interest is on understanding a household's choice behavior simultaneously across the multiple product categories that are represented within the household's shopping basket. We recognize that a single-category choice model provides only a *partial* model of consumer behavior that ignores possible dependencies between the household's purchase outcomes across categories. This leads to a biased understanding of the determinants of consumer choices in the product category. In contrast, a multi-category model (which, in the limit, must include *all* product categories within the household's shopping basket), aims to specify a *full* model of consumer behavior, which provides a better understanding of consumer choices in *any* single product category. Given that researchers working with aggregate data sources have documented cross-category effects in brands' sales both within and across stores (see Walters and MacKenzie, 1988; Walters, 1991; Song and Chintagunta, 2001), it is incumbent upon choice modelers to estimate such effects using disaggregate data sources.

One key reason for previous research ignoring the simultaneous modeling of consumer decisions across multiple product categories is the computational burden associated with estimating such multi-category models. Given recent developments in computational statistics, along with the availability of high-dimensional basket data from research vendors such as IRI, the time is now ripe for choice modelers to undertake multi-category choice modeling. This paper presents a review of where we are, and where we should be going. We present an overview of multi-category models—in terms of (1) the purchase outcomes being modeled, as well as (2) the sections in this paper dealing with each subset of outcomes—in Table 1. Section 6 deals with store choice models that require a multi-category purchasing model as an input. Section 7 concludes.

Table 1. Overview of multi-category models of consumer purchasing.

Consumer purchasing decisions		Section in paper
	One Outcome	
Incidence	Whether to Buy	2.1
	When to Buy	2.2
	Bundle Choice	2.3
Brand Choice	Correlated Marketing Mix Sensitivities	3.1
	Correlated Brand Preferences	3.2
Quantity		4
	Two Outcomes	
Incidence and Brand Choice		5.1
Incidence and Quantity		5.2
	Three Outcomes	
Incidence and Brand Choice and Quantity		5.3

2. Models of Incidence Outcomes in Multiple Categories

2.1. Multi-Category "Whether to Buy" Models

A household's decision of whether or not to buy cake mix on a given shopping trip may not be independent of the household's decision of whether or not to buy frosting on the same shopping trip. In other words, a household's incidence decisions may be related across product categories both because product categories serve as *complements* (e.g., cake mix and frosting) or *substitutes* (e.g., cola and orange juice) in addressing the household's consumption needs, and because product categories vie with each other in attracting the household's limited shopping budget. Manchanda et al., (1999), as well as Chib et al., (2002), employ the panel data multivariate probit (MVP) model to explain household-level contemporaneous incidence outcomes in multiple product categories.

Manchanda et al. (1999) derive their MVP model by assuming that households' indirect utilities for various product categories follow a *joint normal* distribution (instead of *independent normal* distributions, as implied by the binary probit, single-category model). The authors make a substantive distinction between the estimated correlations and the estimated coefficients associated with the cross-price variables (e.g., the coefficient of frosting's price in the household's indirect utility function for cake mix), calling the former *coincidence* and the latter *complementarity*. The authors estimate significant cross-category effects of prices and promotions. While Manchanda et al. (1999) fit their MVP model to households' incidence outcomes in *two* product categories at a time (such as cake mix and frosting), Chib et al. (2002) fit their MVP model simultaneously to households' incidence outcomes in *twelve* product categories. Chib et al. (2002) find that a two-dimensional MVP model, such as that used by Manchanda et al. (1999), underestimates the magnitude of cross-category correlations and overestimates the effectiveness of the marketing mix (i.e., price, display and feature), when compared to the twelve-dimensional MVP model. This underscores the importance of simultaneously modeling a household's incidence outcomes in a large

number of product categories. Chib et al. (2002) uncover high cross-category complementarity (in the form of pair-wise correlations) for the following pairs of product categories: (cola and non-cola), (hot dogs and bacon), (tissue and detergents). The authors also find that ignoring the effects of unobserved heterogeneity across households leads one to overestimate cross-category correlations and underestimate the effectiveness of the marketing mix.

One of the empirical findings in Chib et al. (2002) is that cross-category correlations among all possible pairs among the twelve product categories are *always positive*, if they are different from zero. In other words, there is a base level of complementarity estimated among all pairs of product categories. One reason for this could be the large number of no purchase outcomes that characterizes any product category in scanner panel data. For example, among the 39276 shopping trip observations used by Chib et al. (2002), only 5922 result in the purchase of non-cola beverages, only 5534 result in the purchase of toilet tissue etc. Therefore, since any pair of product categories would share a large number of "zero" (i.e., no purchase) outcomes, they may show up spuriously as complements in the estimation. In order to explicitly address this point, Ma and Seetharaman (2004a) employ the multivariate logit (MVL) model (first proposed by Cox, 1972) to explain households' incidence outcomes across six product categories. Unlike the MVP, the MVL estimates cross-category correlations based on joint purchase outcomes only. Ma and Seetharaman (2004a) find that the estimated cross-category correlations in the joint purchase outcomes from the MVL are all still positive, albeit being smaller in magnitude compared to the estimated correlations from the MVP used by Chib et al. (2002). This suggests that notwithstanding the low frequency of consumer purchasing in most packaged goods categories, there appears to be an intrinsic propensity for any pair of product categories to co-occur within a household's shopping basket. However, to the extent that the estimated correlations are found to be quite different in magnitude across different pairs of product categories, one can still separate strong complements from weak complements using the data. Russell and Peterson (2000) also apply the MVL to explain consumer purchasing among (four) product categories.

2.2. Multi-Category "When to Buy" Models

An alternative model of households' incidence outcomes within a single category, in contrast to a discrete choice model (such as binary logit or probit), is the hazard model. The key feature of the hazard model is that it views a household's inter-purchase time (as opposed to whether or not a household will buy on a given shopping trip) as the outcome variable to be modeled. With multi-category incidence outcomes, therefore, an appropriate alternative to the MVP and MVL models is the Multivariate Hazard model, which would accommodate correlations in a household's inter-purchase times across multiple product categories. Chintagunta and Haldar (1998) model households' joint purchasing behavior for washing machines and dryers using a bivariate hazard model. Since their model admits only positive correlations between the two timing outcomes, it is applicable only to complementary pairs of product categories. A Multivariate Proportional Hazard Model (MVPHM), recently

proposed by Van de Berg (2000), is employed by Ma and Seetharaman (2004b) to explain households' incidence outcomes across six product categories while admitting both positive and negative pair-wise correlations in the outcomes (for alternative specifications of the MVPHM, see Hougaard, 2000; Pipper and Martinussen, 2004; Bhat et al., 2004). Additionally the authors also estimate a multivariate version of the additive risk model (ARM), referred to as MVARM, whose univariate version has been shown to outperform the PHM in predicting single-category incidence outcomes (Seetharaman, 2004). The authors find that the MVARM fits incidence outcomes better than the MVPHM, which in turn fits the outcomes better than existing univariate hazard models such as the PHM and ARM. Importantly, the estimates from the MVARM yield non-monotonic time-varying price elasticities of demand for all product categories under study (while the MVP or the MVL would restrict price elasticities of demand to be time invariant).

2.3. "Bundle Choice" Models

Another consumer-level model of multi-category incidence outcomes is the bundle choice model (BCM). This model explains how a consumer decides whether or not to buy a *bundle* of product categories—such as personal computer, monitor and printer—at a given shopping occasion on the basis of the product attributes that are embodied in the product categories. One example of a BCM is the model of Chung and Rao (2003), that is built on the premise that consumers classify all the relevant attributes into three types according to the degree of comparability among product categories: (i) fully comparable attributes, i.e., ubiquitous attributes that exist in all product categories of the bundle (e.g., brand reliability), (ii) partially comparable attributes, i.e., attributes that are applicable to more than one but not all of the product categories of the bundle (e.g., speed of processing, which applies to the computer and printer, but not to the monitor), and (iii) non-comparable attributes, i.e., attributes unique to a single product category in the bundle (e.g., screen size of monitor). The BCM also allows these attributes to be of the following two types: (i) *non-balancing* attributes, i.e., attributes whose average presence in the bundle should be either high or low for the consumer, (ii) *balancing* attributes, i.e., attributes whose dispersion among different products in the bundle should be either high or low for the consumer. The error terms associated with the consumer's utilities for the various possible bundles and the no purchase option are assumed to follow a joint gumbel distribution in such a manner so as to yield a nested logit model of bundle choice at the consumer level. The authors estimate this nested logit model using experimental choice data collected from respondents (where each choice task involves the respondent choosing among a set of bundles and the no purchase option). The authors show how to use their proposed experimental approach to (1) find market segments for bundles with heterogeneous products in multiple categories, (2) estimate individual reservation prices for bundles, and (2) determine optimal bundle prices for different market segments.

A second example of a BCM is the model of Jedidi et al. (2003) that defines the consumer's (random) utility for a bundle as the sum of the consumer's reservation price for the bundle plus a random component (that the authors motivate as capturing consumer "errors

in judgment") that is distributed independent normal (with different variances) across bundles. This yields multinomial probit probabilities for the consumer's choice among possible bundles (and the no purchase decision). The authors estimate their BCM using experimental data on respondents' choices in each of three product groups: (i) video camera and videocassette player, (ii) six-month subscriptions to Time and Newsweek magazines, and (iii) microwave oven and color television. The authors derive implications for product-line pricing, and find that a uniformly high price for all products and bundles is optimal when the degree of heterogeneity in reservation prices across consumers is high, and a hybrid strategy is optimal otherwise.

What distinguishes the BCM from the Multivariate Discrete Choice Models (e.g., MVP, MVL) and the Multivariate Hazard Models (e.g., MVPHM, MVARM) is that the BCM operationalizes consumer choices directly at the level of the product bundle, instead of at the level of the product categories. An additional feature of the BCM of Chung and Rao (2003) is that it is specified in terms of product attributes encapsulated in the bundle (see Rao, (2004) for a discussion of alternative attribute-based operationalizations of the consumer's bundle choice problem). Another related stream of literature on BCM is based on collecting conjoint data from respondents for bundles of products, and then estimating both the main effects for each product and the interaction effects between products in consumers' utility functions (Green et al., 1972; Green and Devita, 1974; Goldberg et al., 1984).

2.4. Directions for Future Research

The appeal of the MVP model, discussed in Section 2.1, lies in the flexibility of the cross-category relationships that can be estimated using basket data. The appeal of the MVPHM and MVARM models, discussed in Sections 2.2, lies in their ability to accommodate the effects of time-varying price elasticities in consumer purchasing behavior. The appeal of the BCM model, discussed in Section 2.3, lies in its ability to parsimoniously describe multi-category purchasing behavior of consumers using underlying product attributes. It would be useful to compare the explanatory and predictive abilities of the three modeling approaches to see which modeling aspect is more important than the others. If no clear winner emerges out of such a comparison, it would then be useful to develop a unified multi-category incidence model that simultaneously exploits the good features—flexibility, time-varying responsiveness and parsimony—of the three approaches.

3. Brand Choice Outcome Models in Multiple Categories

3.1. Correlated Marketing Mix Sensitivities Across Categories

Ainslie and Rossi (1998) investigate whether a household's brand choice outcomes in multiple product categories are correlated on account of the household having common sensitivities to the marketing mix (i.e., price, display and feature) across product categories. Using a variance components decomposition—into (i) a household-specific component and (ii) a category-specific component—of the marketing mix coefficients in a household's

utility function for brands (that yields a Multinomial Probit (MNP) model of brand choices) within each of five product categories, the authors estimate significant correlations in household responsiveness to price (0.32) and feature advertising (0.58) across product categories. An important precursor to Ainslie and Rossi (1998) is Kim et al. (1999), who investigate the same research question as Ainslie and Rossi (1998), but by first estimating single-category brand choice models separately for five product categories, and then computing pair-wise correlations between the five sets of point estimates of model parameters. Ainslie and Rossi (1998) demonstrate that such a two-stage approach underestimates cross-category correlations in model parameters. Seetharaman et al. (1999) extend the Ainslie and Rossi (1998) model to investigate whether a household's state dependence behavior (i.e., propensity to seek inertia or variety in its brand choices over time) is correlated across product categories. They conclude that if a household is strongly inertial (i.e., buys the same brand over time "out of habit") in one product category, then the household is also strongly inertial in other product categories.

Iyengar et al. (2003) look at the situation of having limited or no information about a consumer's purchasing behavior in a product category, investigating whether one could leverage information from the purchases of this consumer in other product categories (where their purchases are fully observed) that are related to the focal product category. They lay out the conditions under which such leveraging is indeed fruitful, and contrast them with the conditions under which it is better to leverage information from the purchases of other consumers in the focal product category. Similar to Ainslie and Rossi (1998) and Seetharaman et al. (1999), the authors uncover a high correlation (0.21) in households' price coefficients across categories. They find that leveraging information across categories is most useful when there is complete information on a consumer in one category, but no or very little information about the consumer in a related category. They also find that such leveraging is not useful when firms have no causal information (i.e., marketing variables) in either category.

3.2. Correlated Brand Preferences Across Categories

Russell and Kamakura (1997) model inter-category correlations in households' purchase volumes of brands (aggregated over time) in four product categories. Using a Poisson model for brands' purchase volumes, and panel data from 626 Canadian households, the authors find that consumer preferences for the store brand name are consistent (i.e., positively correlated) across categories, while such cross-category consistency is not observed for all national brands.

Erdem (1998) and Erdem and Winer (1999) study if consumers' quality perceptions of brands are correlated across two product categories on account of these brands sharing a common brand name. Such a correlation would be predicted by the signaling theory of *umbrella branding*. Using a Multinomial Logit (MNL) brand choice model that incorporates consumer learning under product quality uncertainty, the authors finds strong empirical evidence for such cross-category correlations within a given brand name. Further, the authors uncover negative correlations between different brand names across categories

(e.g., between Crest toothpaste and Colgate toothbrush). Erdem and Sun (2002) extend these models to show that cross-category advertising and sales promotion spill-over effects also exist for such umbrella brands.

Singh et al. (2005) estimate brand choice models simultaneously across three closely related, and two less closely related, product categories for a set of 1017 households observed over a two-year period. The model allows not only the marketing mix coefficients (as in Ainslie and Rossi, 1998; Seetharaman et al., 1999) in the consumer's utility function for brands (that yields an MNL model of brand choices), but also the consumer's brand preferences, to be correlated across product categories, using a factor-analytic structure. Singh et al. (2005) uncover high correlations in households' preference for private labels (0.40), for a national brand (0.34), as well for an attribute called "fat-free " (0.79), across product categories. The authors use their estimation results to predict customer behavior for a new product not only in one of the product categories under study, but also in a new product category (that shares an attribute with the product categories under study) on which no purchase data is available.

Another paper that estimates cross-category correlations in consumers' preferences for private labels is Ainslie et al. (2004). Using a random utility specification (that yields the MNL model of brand choices) that restricts the consumer's price parameter to be -1 (in order to be able to estimate the scale parameter associated with the consumer's utility function), the authors apply the variance components approach of Ainslie and Rossi (1998) and Seetharaman et al. (1999) to estimate cross-category correlations in not only the utility parameters, but also the scale parameter. Consistent with Singh et al. (2005), the authors uncover high cross-category correlations in private label preferences. Further, the authors find that more educated consumers, as well as lower income consumers, show higher private label preferences than other consumers. Interestingly, the authors also find that a benchmark random utility model that restricts the scale parameter to be 1 and, instead, estimates the price parameter (not fixing it at -1), is not able to recover these high correlations.

A fifth paper that investigates cross-category correlations in households' private label preferences is Hansen et al. (2003). Like Singh et al. (2005), the authors use a factor structure to obtain a parsimonious representation of these cross-category correlations between MNL models of brand choices in ten product categories. The authors uncover strong cross-category correlations in households' private label preferences. These correlations are found to be stronger for non-food categories than for food categories.

3.3. Directions for Future Research

The experimental work of Dhar and Simonson (1999) can be used to add useful behavioral insights to the existing empirical research on multi-category brand choice behavior. The authors investigate, using a series of experimental studies in the laboratory, consumption episode effects on how consumers make contemporaneous brand choices in different product categories. The authors find that (i) in episodes involving a tradeoff between a goal (e.g., pleasure) and a resource (e.g., money), consumers tend to highlight either goal fulfillment or resource conservation by selecting similar attribute levels for products consumed in the

same episode (e.g., a tasty, expensive appetizer and a tasty expensive entree on one occasion, versus less tasty, less expensive items on another occasion), and (ii) if each choice involves a tradeoff between two goals (e.g., pleasure and good health), consumers tend to balance attribute levels (e.g., in each episode, have one tasty item and one healthy item). While scanner panel data do not record consumer tradeoffs between goals and resources, it would be useful to investigate whether (i) consumers make consistent attribute choices in some shopping trips, and balanced attribute choices in others, (ii) consumer demographics (which are observed) are systematically correlated with whether consumers generally balance attribute levels or not across product categories.

Multi-category brand choice models can be developed to understand consumers' brand choice behavior across disparate product categories (e.g. homes, automobiles, restaurants etc.) Given the availability of detailed customer behavioral lists from list brokers, such analyses may now be practically feasible and will aid firms in different industries (say, Toyota Camry and Olive Garden) to jointly develop cross-promotional and cross-selling strategies for their brands. Differences between *contemporaneous* multi-category buying (observed in supermarkets) and *sequential* multi-category buying (observed for durable goods and services), are also worth studying.

4. Models of Quantity Outcomes in Multiple Categories

Borle et al. (2004) employ the Dirichlet distribution to model the household's expenditure allocation across 147 product categories (taking the household's total shopping expenditure as given), where the Dirichlet parameters capture the relative shares of the product categories within the household's shopping basket. However, there is no existing work on consumer-level multi-category models of quantity outcomes. This is surprising since in some product categories—paper towels, toilet tissue, canned tuna etc.—consumers engage in substantively significant multi-unit buying. However, quantity outcomes have been modeled in the context of also simultaneously modeling incidence and/or brand choice outcomes, as will be discussed in Sections 5.2 and 5.3. It would be of interest to investigate whether modeling cross-category relationships in quantity outcomes only, while ignoring incidence and brand choice outcomes, provides similar substantive insights to those obtained from the models discussed in Sections 5.2 and 5.3.

5. Models of Multiple Outcomes in Multiple Categories

5.1. Incidence and Brand Choice

5.1.1. Treating Incidence as an Alternative in a Multinomial Choice Model Deepak et al. (2002) extend the work of Ainslie and Rossi (1998) to investigate whether households' price sensitivities are correlated across product categories using a Multivariate Probit (MVP) model of incidence and brand choice outcomes. Cross-category correlations in incidence outcomes are modeled as in Manchanda et al. (1999) (through correlations between the error terms of the consumer's indirect utility functions for different product categories), while

cross-category correlations in marketing mix sensitivities are accommodated in a flexible manner. Consistent with Ainslie and Rossi (1998) and Seetharaman et al. (1999), the authors uncover significant cross-category correlations in marketing mix sensitivities of households.

Ma, Seetharaman and Narasimhan (2005) model incidence and brand choice outcomes of households across two product categories using a Multivariate Logit (MVL) model. Further, the authors look at the pricing implications of the estimated cross-category effects for manufacturers (including those, such as General Mills, who sell *product lines*, rather than single products, within either product category) and retailers. For this purpose, the authors employ recently developed empirical industrial organization techniques (Sudhir, 2001; Villas-Boas and Zhao, 2005) and find that (i) single-category models of incidence and brand choice overestimate price elasticities of demand for the brands in the two categories, (ii) observed prices in the cake mix and frosting categories are consistent with (a) category profit maximization on the part of the retailer, (b) vertical Nash interactions between the manufacturers and the retailer, and (c) product mix profit maximization on the part of each manufacturer.

5.1.2. Treating Incidence and Brand Choice as Distinct Decision Stages Mehta (2005) derives a simultaneous model of incidence and brand choice outcomes in multiple (i.e., C) product categories as the first order conditions of a consumer's basket utility maximization problem. The author estimates the proposed model using consumer purchasing data on three product categories, and finds that (i) washer softener and dryer softener are weak purchase complements, (ii) liquid detergent and washer softener show strong cross-price effects, and (iii) consumers show similar price sensitivity between liquid detergent and dryer softener.

Chib et al. (2005) extend the multi-category incidence model of Chib et al. (2002) to additionally model brand choice outcomes within each product category. This is done by first invoking a single-category model of brand choice with a no purchase option developed by Chib et al. (2004), for one product category at a time, and then overlaying cross-category correlations in the no-purchase outcomes as in Chib et al. (2002). In addition, households' indirect utilities for brands are also allowed to be correlated across product categories in a flexible manner. The authors estimate effects of "umbrella branding" strategies adopted by private labels (e.g., President's Choice), insofar as they lead to correlated consumer preferences for the same brand name in different product categories.

5.2. Incidence and Quantity

Niraj et al. (2004) estimate a two-stage bivariate logit model of incidence and quantity outcomes of households across three pairs of product categories. Their empirical results demonstrate that promotional spillovers vary systematically across the three pairs. The category pair of bacon and eggs show cross-category correlations in both incidence and quantity, while the category pair of detergent and softener show cross-category correlations in incidence only, and the category pair of detergent and analgesics do not show cross-category correlations in either decision. The authors decompose the cross-category impact of promotions into incidence and quantity components, and find that incidence effects account for 60%.

5.3. Incidence, Brand Choice and Quantity

Song and Chintagunta (2004) estimate a simultaneous model of the three purchase outcomes—incidence, brand choice and quantity—in two product categories. The proposed multi-category model is derived using a utility maximizing framework, under which the consumer maximizes, subject to a budget constraint, a direct utility function that permits interior solutions on incidence (i.e., purchase of more than one product category at the same time), but allows only for corner solutions on brand choice. The authors find that the majority of the cross-category effects arise in the incidence and brand choice outcomes, rather than in the quantity outcomes.

5.4. Directions for Future Research

While previous research has simultaneously modeled incidence and brand choice (Section 5.1), or incidence and quantity (Section 5.2), a simultaneous model of brand choice and quantity has not been estimated so far. Estimating such a model, and comparing it to the model of Song and Chintagunta (2004) (Section 5.3), would reveal the nature of biases that arise from ignoring incidence outcomes in the estimation. If such a comparison shows that ignoring incidence outcomes is not expensive (in terms of the resulting bias in parameter estimates), one can achieve enormous computational gains from working with smaller datasets that ignore no purchase outcomes (which constitute 90% of observations in most scanner panel datasets).

It would also be useful to extend the flexible multi-category model of incidence and brand choice proposed by Chib et al. (2005) to additionally handle quantity outcomes, and then compare its empirical performance to that of Song and Chintagunta (2004). This will reveal the value (or lack thereof) of imposing theory-based restrictions on the estimable statistical model of incidence, brand choice and quantity outcomes.

5.5. Estimation Issues

Multi-category models are computationally burdensome in terms of the number of parameters to be estimated. Frequentist approaches to estimation would typically fail to handle this estimation burden unless one focuses on a limited (i.e., two or three) number of product categories and imposes theory-based restrictions on the estimable set of parameters, for example, as in Mehta (2005). However, if one's interest, instead, is in estimating a flexible statistical model of purchase outcomes without imposing parameter restrictions (since there is little historical research wisdom on multi-category choice behavior of households to yield convenient *a priori* restrictions), Markov Chain Monte Carlo (MCMC) methods have to be invoked to enable the estimation of these models in the *Bayesian* realm. A seminal paper that lays the blueprint for such estimation in these modeling contexts is Albert and Chib (1993). The Albert and Chib approach is based on the tactical introduction of latent variables as additional unknowns to simplify the analysis of the posterior distribution by MCMC methods. The Albert and Chib technique, as formulated in Chib

and Greenberg (1998), serves as the basic building block in developing MCMC samplers for multi-category choice models. What underlies the MCMC estimation are (1) the *Bayes Theorem*, that says that sampling the posterior density of the model parameters must be the focus of estimation, and (2) the idea that one can use conditional distributions of model parameters to obtain their joint distribution. Generating correlated draws from a Markov chain—whose limiting, invariant distribution is the target distribution - is the central idea behind the estimation. Two popular MCMC algorithms that have been applied to estimate multi-category models are the *Gibbs Sampler* and the *Metropolis-Hastings algorithm* (see Chib and Greenberg, 1995). For applications, see Manchanda et al. (1999), Chib et al. (2002, 2005).

6. Models of Store Choice Outcomes

Bell and Lattin (1998) model households' store choice outcomes using an MNL model, with two types of explanatory variables at the store-level: (i) expected basket attractiveness (i.e., composite attractiveness of a basket of product categories at the store), and (ii) geographic distance of consumer to the store. The expected basket attractiveness is defined as a composite of the expected attractiveness of all product categories within the household's shopping basket. The expected attractiveness of a product category, in turn, is taken to be the estimated inclusive value from a nested logit model of incidence and brand choice outcomes in the product category. Households' incidence decisions across product categories are assumed, however, to be independent. The authors find that expected basket attractiveness has a significant effect on households' store choice decisions, and that large basket shoppers have larger store-level elasticities.

Bell et al. (1998) disentangle the effects of two types of costs—(1) fixed costs, i.e., the household's store loyalty and distance to the store, and (2) variable costs, i.e., expected costs of the household's shopping list—on households' store choice behavior using an MNL model. The authors find both types of costs to be significant, and that EDLP stores impose higher average fixed costs of shopping compared to HiLo stores. The authors also estimate threshold basket sizes at which households are indifferent between visiting a nearby store with lower (higher) fixed (variable) cost, and a farther store with higher (lower) fixed (variable) cost.

Bodapati and Srinivasan (2001) model households' store choice outcomes using a nested MNL model, where the first stage involves the consumer's store choice decision, the second stage involves multi-category incidence decisions at the chosen store, and the third stage involves brand choice decisions within the chosen product categories. Unlike Bell and Lattin (1998)—who estimate the store choice model separately from the nested logit model of incidence and brand choice—the authors *jointly* estimate the three decisions using MCMC methods. Further, the authors also allow a household's incidence decisions to be correlated across categories (using an MVP model). The key finding is that newspaper feature advertising plays an important role in influencing store choice decisions of households, since such advertising serves as the basis for households forming price expectations about products prior to undertaking shopping trips.

Chan et al. (2005) investigate the drivers of consumers' store choice decisions using AC Nielsen scanner panel data from 642 Denver area consumers in 37 product categories during a 119-week period spanning January 1993–April 1995. Since the consumer purchase data are collected using home scanners, the coverage of stores is quite exhaustive, i.e., the 256 retailers in the data include convenience stores and mass merchandisers in addition to supermarkets. Stores' geographic locations, as well as consumers' geographic locations are known. Using data only on major shopping trips (i.e., trips that involve a pre-defined minimum expenditure), which reduces the number of stores used in the estimation to 47, the authors explain consumers' store choice outcomes using an MNL model with two types of explanatory variables at the store-level: (i) basket attractiveness, and (ii) geographic distance of consumer to the store. Basket attractiveness is defined as the sum of attractiveness of all product categories under consideration, where each product category's attractiveness is defined as the attractiveness of the most attractive brand in the product category, which is identified on the basis of a brand choice model estimated separately for each product category. The product category attractiveness also takes into account the household's expected number of units of the product to be purchased, which is estimated using a consumption needs model (where seasonality is allowed to influence whether or not a product category is likely to be needed during a shopping trip, and the conditional number of units to be purchased is taken to be a gaussian random variable whose mean and variance are estimated using the household's purchase quantities in the data).

Borle et al. (2004) estimate the effects of cuts in product assortments—where the number of SKUs eliminated varies from 24% to 91%, and comprise mostly low share items—in 147 product categories at an online grocer using a test group of 840 customers (who experienced the assortment cut) versus a control group of 378 customers (who experienced no cut in assortment). Based on analyzing purchasing data from the two groups of customers during a six-month time window prior to the assortment cut, and a six-month time window after the assortment cut, the authors find that inter-shopping times of customers (which are modeled using the COM-Poisson distribution, developed by Boatwright et al., 2003) increased by 25%, while purchase amounts (which are modeled using a log-normal distribution) decreased by 5%, after the assortment cut. This suggests that there may be adverse consequences for store patronage and store traffic from storewide cuts in product assortments. The authors suggest that managers must focus their cuts, therefore, in select categories.

6.1. Directions for Future Research

It would be of interest to understand the drivers of observed *portfolios* of stores where consumers shop. For example, a consumer may buy produce at the supermarket, milk and soda at a convenience store, and frozen entrees at a discount store, all within the same week. A model that ignores such multi-format shopping of consumers, and instead treats each store visit as an independent discrete choice among a set of stores (as is commonly done in existing work), would be mis-specified in terms of explaining observed store patronage behavior of consumers. Application of newer statistical methods, such as Bayesian networks, could reduce the computational burden associated with estimating such models.

7. Conclusions

In this paper, we review multi-category models in terms of four types of purchase outcomes: incidence, brand choice, quantity and store choice. The majority of the work in this area is ongoing, which indicates its contemporary relevance and importance. Given that the past three Choice Symposia have devoted sessions to the theme of multi-category choice modeling, we cannot under-emphasize the continuing importance of this topic within our academic discipline. Renewed research activity on multi-category issues has been spurred by two recent developments: (i) the availability of rich market basket data, (ii) the development of computational techniques that enables the estimation of high-dimensional multi-category models. With the emergence of even more comprehensive databases, such as a recently constructed dataset at SUNY-Buffalo, we believe that this research activity will gain more momentum in the years to come. Researchers hoping to work in this area can use this review paper to time-effectively gain an understanding of the prevailing wisdom on multi-category issues, as well as identify and correct possible "gaps" in current research.

Acknowledgment

We thank David Bell, Sharad Borle and Jeongwen Chiang for their comments.

References

Ainslie, A. and P. E. Rossi. (1998). "Similarities in Choice Behavior Across Multiple Categories," *Marketing Science* 17(2), 91–106.

Ainslie, A., G. Sonnier, and S. Moorthy. (2004). "Taming the Intercept Term in Choice Models: An Application to Private Labels," Working Paper, University of California at Los Angeles.

Albert, J. and S. Chib. (1993). "Bayesian Analysis of Binary and Polychotomous Response Data," *Journal of the American Statistical Association* 88, 669–679.

Bell, D. R. and J. M. Lattin. (1998). "Shopping Behavior and Consumer Preferences for Store Price Format: Why Large Basket Shoppers Prefer EDLP," *Marketing Science* 17(1), 66–88.

Bell, D. R., T. Ho, and C. S. Tang. (1998). "Determining Where to Shop: Fixed and Variable Costs of Shopping," *Journal of Marketing Research* 35(3), 352–369.

Bhat, C. R., S. Srinivasan, and K. W. Axhausen. (2004). "An Analysis of Multiple Interactivity Durations Using a Unifying Multivariate Hazard Model," Working Paper, The University of Texas at Austin.

Bodapati, A. V. and V. Srinivasan. (2001). "The Impact of Out-of-Store Advertising on Store Sales, "Working Paper," University of California at Los Angeles.

Boatwright, P., S. Borle, and J. B. Kadane. (2003). "A Model of the Joint Distribution of Purchase Quantity and Timing," *Journal of the American Statistical Association* 98, 564–572.

Borle, S., P. Boatwright, J. B. Kadane, J. C. Nunes, and G. Shmueli. (2004). "Effect of Product Assortment Changes on Customer Retention," Working Paper, Carnegie Mellon University.

Chiang, J. (1991). "A Simultaneous Approach to the Whether, What and How Much to Buy Questions," *Marketing Science* 10(4), 297–315.

Chan, T., Y. Ma, C. Narasimhan, and V. Singh. (2005). "A Store Choice Model with Basket Shopping," Working Paper, Washington University in St. Louis.

Chib, S. and E. Greenberg. (1995). "Understanding the Metropolis-Hastings Algorithm," *The American Statistician* 49(4), 327–335.

Chib, S. and E. Greenberg. (1998). "Analysis of Multivariate Probit Models," *Biometrika* 85(2), 347–361.

Chib, S., P. B. Seetharaman, and A. Strijnev. (2002). "Analysis of Multi-Category Purchase Incidence Decisions Using IRI Market Basket Data," *Advances in Econometrics* 16, 55–90.

Chib, S., P. B. Seetharaman, and A. Strijnev. (2004). "Model of Brand Choice With a No-Purchase Option Calibrated to Scanner-Panel Data," *Journal of Marketing Research* 41(2), 184–196.

Chib, S., P. B. Seetharaman, and A. Strijnev. (2005). "Joint Modeling of Multiple Category Incidence and Private Label Choice, Accounting for No Category Incidence and Panel Heterogeneity," Working Paper, School of Management, University at Buffalo.

Chintagunta, P. K. and S. Haldar. (1998). "Investigating Purchase Timing Behavior in Two Related Product Categories," *Journal of Marketing Research* 35(1), 43–53.

Chung, J. and V. R. Rao. (2003). "A General Choice Model for Bundles with Multiple-Category Products: Application to Market Segmentation and Optimal Pricing for Bundles," *Journal of Marketing Research* 40(2), 115–130.

Cox, D. R. (1972). "The Analysis of Multivariate Binary Data, Applied Statistics," *Journal of the Royal Statistical Society, Series C* 21(2), 113–120.

Deepak, S., A. Ansari, and S. Gupta. (2002). "Investigating Consumer Price Sensitivities Across Categories," Working Paper, University of Iowa.

Dhar, R. and I. Simonson. (1999). "Making Complementary Choices in Consumption Episodes: Highlighting Versus Balancing," *Journal of Marketing Research* 36(1), 29–44.

Elrod, T., G. J. Russell, A. D. Shocker, R. L. Andrews, L. Bacon, B. L. Bayus, J. D. Carroll, R. M. Johnson, W. A. Kamakura, P. Lenk, J. A. Mazanec, V. R. Rao, and V. Shankar. (2002). "Inferring Market Structure from Customer Response to Competing and Complementary Products," *Marketing Letters* 13(3) 221–232.

Erdem, T. (1998). "An Empirical Analysis of Umbrella Branding," *Journal of Marketing Research* 35(3), 339–351.

Erdem, T. and B. Sun. (2002). "An Empirical Investigation of the Spillover Effects of Advertising and Sales Promotions in Umbrella Branding," *Journal of Marketing Research* 39, 1–16.

Erdem, T. and R. Winer. (1999). "Econometric Modeling of Competition: A Multicategory Choice-Based Mapping Approach," *Journal of Econometrics* 89, 159–175.

Goldberg, S. M., P. E. Green, and Y. Wind. (1984). "Conjoint Analysis of Price Premiums for Hotel Amenities," *Journal of Business* 54(1), 111–132.

Green, P. E. and M. T. Devita. (1974). "A Complementarity Model of Consumer Utility for Item Collections," *Journal of Consumer Research* 1(4), 56–67.

Green, P. E., Y. Wind, and A. K. Jain. (1972). "Preference Measurement of Item Collections," *Journal of Marketing Research* 9(4), 371–377.

Hansen, K., V. Singh, and P. K. Chintagunta. (2003). "Understanding Store Brand Purchase Behavior Across Categories," Working Paper, Northwestern University.

Ho, T., C. S. Tang, and D. R. Bell. (1998). "Rational Shopping Behavior and the Option Value of Variable Pricing," *Management Science* 44(12), 145–160.

Hougaard, P. (2000). *Analysis of Multivariate Survival Data*. New York: Springer Verlag.

Iyengar, R., A. Ansari, and S. Gupta. (2003). "Leveraging Information Across Categories," *Quantitative Marketing and Economics* 1(4), 425–465.

Jedidi, K., S. Jagpal, and P. Manchanda. (2003). "Measuring Heterogeneous Reservation Prices for Product Bundles," *Marketing Science*. 22(1), 107–130.

Kamakura, W. A., S. Ramaswami, and R. K. Srivastava. (1991). "Qualification of Prospects for Cross-Selling in the Financial Services Industry," *International Journal of Research in Marketing* 26(4), 379–390.

Kim, B. D., K. Srinivasan, and R. T. Wilcox. (1999). "Identifying Price Sensitive Consumers: The Relative Merits of Demographic versus Purchase Pattern Information," *Journal of Retailing* 75(2), 173–193.

Ma, Y., P. B. Seetharaman, and C. Narasimhan. (2005). "Empirical Analysis of Competitive Pricing Strategies with Complementary Product Lines," Working Paper, Washington University in St. Louis.

Ma, Y. and P. B. Seetharaman. (2004a). "The Multivariate Logit Model for Multicategory Purchase Incidence Outcomes," Working Paper, Rice University.

Ma, Y. and P. B. Seetharaman. (2004b). "Multivariate Hazard Models for Multicategory Purchase Timing Behavior," Working Paper, Rice University.

Manchanda, P., A. Ansari, and S. Gupta. (1999). "The Shopping Basket: A Model for Multicategory Purchase Incidence Decisions," *Marketing Science* 18(2), 95–114.

Mehta, N. (2005). "Investigating Consumers' Purchase Incidence and Brand Choice Decisions Across Multiple Product Categories," Working Paper, University of Toronto.

Niraj, R., V. Padmanabhan, and P. B. Seetharaman (2002). "A Cross-Category Model of Households' Incidence and Quantity Decisions," Working paper, University of Southern California.

Pipper, C. B. and T. Martinussen. (2004). "An Estimating Equation for Parametric Shared Frailty Models with Marginal Additive Hazards," Working Paper, The Royal Veterinary and Agricultural University, Frederiksberg C, Denmark.

Rao, V. R. (2004). "Bundles of Multi-Attributed Items: Modeling Perceptions, Preferences and Choices," Working Paper, Cornell University.

Russell, G. J. and W. A. Kamakura. (1997). "Modeling Multiple Category Brand Preference with Household Basket Data," *Journal of Retailing* 73(1), 439–461.

Russell, G. J. and A. Peterson. (2000). "Analysis of Cross Category Dependence in Market Basket Selection," *Journal of Retailing* 76(3), 367–392.

Russell, G. J., D. R. Bell, A. Bodapati, C. Brown, J. W. Chiang, G. Gaeth, S. Gupta, and P. Manchanda. (1997). "Perspectives on Multiple Category Choice," *Marketing Letters* 8(3), 297–305.

Russell, G. J., S. Ratneshwar, A. D. Shocker, D. R. Bell, A. Bodapati, A. Degeratu, L. Hildebrandt, N. Kim, S. Ramaswami, and V. Shankar. (1999). "Multiple Category Decision Making: Review and Synthesis," *Marketing Letters* 10(3), 319–332.

Seetharaman, P. B. (2004). "The Additive Risk Model for Purchase Timing," *Marketing Science* 23(2), 234–242.

Seetharaman, P. B., A. Ainslie, and P. K. Chintagunta (1999). "Investigating Household State Dependence Effects Across Categories," *Journal of Marketing Research* 36(4), 488–500.

Singh, V. P., K. Hansen, and S. Gupta. (2005). "Modeling Preferences For Common Attributes in Multicategory Brand Choice," *Journal of Marketing Research* 42(2), 195–209.

Song, I. and P. K. Chintagunta. (2001). "Investigating Cross-Category Effects of Retailer's Marketing Activities: Application of Random Coefficient Choice Models with Aggregate Data," Working Paper, University of Chicago.

Song, I. and P. K. Chintagunta. (2004). "A Discrete/Continuous Model for Multi-Category Behavior of Households," Working Paper, University of Chicago.

Sudhir, K. (2001). "Structural Analysis of Manufacturer Pricing in the Presence of a Strategic Retailer," *Marketing Science* 20(3), 244–264.

Van den Berg, G. J. (2000). "Duration Models: Specification, Identification, and Multiple Durations," Working Paper, Free University, Amsterdam, The Netherlands.

Villas-Boas, J. M. and Y. Zhao. (2005). "Retailers, Manufacturers, and Individual Consumers: Modeling the Supply Side in the Ketchup Marketplace," *Journal of Marketing Research* 42(1), 83–95.

Walters, R. G. (1991). "Assessing the Impact of Retail Price Promotions on Product Substitution, Complementary Purchase and Inter-store Displacement," *Journal of Marketing* 55(1), 17–28.

Walters, R. G. and S. B. MacKenzie. (1988). "A Structural Equation Model of the Impact of Price Promotions on Store Performance," *Journal of Marketing Research* 25(4), 551–563.