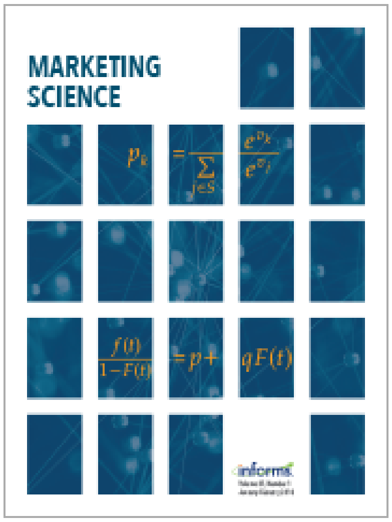


Marketing Science

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

Understanding Store-Brand Purchase Behavior Across Categories

Karsten Hansen, Vishal Singh, Pradeep Chintagunta,

MARKETING SCIENCE
Vol. 25, No. 1, January–February 2006, pp. 75–90
ISSN 0732-2399 | EISSN 1526-548X | 06 | 2501 | 0075

DOI 10.1287/mksc.1050.0151

Understanding Store-Brand Purchase Behavior Across Categories

Karsten Hansen
Department of Marketing, Kellogg School of Management, 2001 Sheridan Road, Northwestern University, Evanston, Illinois 60208, karsten-hansen@northwestern.edu

Vishal Singh
Tepper School of Business, Carnegie Mellon University, 5000 Forbes Avenue, 244 Posner Hall, Pittsburgh, Pennsylvania 15213, vsingh@andrew.cmu.edu

Pradeep Chintagunta
Graduate School of Business, University of Chicago, 5807 South Woodlawn Avenue, Chicago, Illinois 60637, pradeep.chintagunta@chicagogsb.edu

This paper investigates whether the tendency to buy store brand is category specific, or an enduring consumer trait. We develop a multicategory brand-choice model with a factor-analytic structure on the covariance matrix of the coefficients. The methodology allows us to elicit the basic latent tendency for a household to buy store brands, while controlling for other causes such as price sensitivity. The model is applied to a set of ten food and nonfood product categories. We find strong evidence of correlations in household preferences for store brands across categories. Using a two-dimensional factor structure, we find that one of the factors explains a substantial amount of variation in store-brand preference, while the other factor explains price sensitivity—*consistently across categories*. The presence of these factors in all categories indicates that there are unobservable household-level traits that are non-category specific, i.e., stable across product categories. Using data from five holdout categories, we find that household estimates of these latent factors are very useful in predicting demand for store brands in new categories. Other potential applications for store managers are discussed.

Key words: store brands; multicategory choice models; heterogeneity; frequent-shopper data
History: This paper was received December 23, 2003, and was with the authors 9 months for 2 revisions; processed by Michel Wedel.

1. Introduction

One of the most important activities for supermarket retailers is the creation and maintenance of their private-label brands. Store brands are particularly important to retailers, as they are the only brands that are available exclusively at their store. A well-developed private-label program could not only contribute directly to retailer profitability, but also have positive indirect effects such as better bargaining power with the manufacturer (Mills 1995) or building store loyalty (Corstjens and Lal 2000). According to the Private-Label Manufacturers Association (PLMA), store brands now account for one in every five items sold in the U.S. supermarkets and represent a nearly $50 billion segment of the retailing business.

Given their strategic importance, it is not surprising that private labels have generated tremendous interest in the academic literature. Researchers have addressed a wide variety of issues, such as explaining the penetration of private labels across categories (Sethuraman 1992, Hoch and Banerji 1993), penetration of store brands across retailers (Dhar and Hoch 1997), strategic interaction between national brands and private labels (Cotterill et al. 2000), positioning of store brands vis-a-vis national brands (Sayman et al. 2002), conditions for entry of store brand and its impact on retailer's pricing and bargaining power with manufacturers (Raju et al. 1995, Scott-Morton and Zettelmeyer 2000, Chintagunta et al. 2002, Pauwels and Srinivasan 2004), and promotional strategies of national and store brands (Blattberg and Wisniewski 1989, Allenby and Rossi 1991), etc.

From the consumer's perspective, private labels are quite unique in that they are the only brands that recur throughout the store. Indeed, as Hoch (1996) points out, even the largest manufacturers do not come close to the private label in terms of storewide coverage. Given such wide penetration, a natural question that arises is whether consumers exhibit similarities in their preference for store brands across product categories. In other words, is household private-label brand preference category specific or an enduring household trait? Previous research (e.g., Hoch and Banerji 1993) has found significant differences in store-brand penetration across categories, which suggests that category-specific factors could be dominant. At the same time, there are various reasons why preference for private label could be household specific. For instance, if the choice of private label is primarily driven by economic factors such as income constraints, then one might expect lower-income households to buy private label across many categories.[1] A recent stream of literature focusing on general cross-category traits of consumer behavior do find that consumers exhibit some similarities in their purchase behavior across different product categories (Ainslie and Rossi 1998, Erdem 1998, Seetharaman et al. 1999, Singh et al. 2005).

Besides their storewide coverage, another distinguishing feature of private labels is that they are, almost always, the cheapest brand in a product category. This complicates the study of fundamental drivers of store-brand choice. For example, a very simple approach to measuring a household's store-brand proneness is to use the household's observed choice frequency of store brand across categories (over a certain time period).[2] However, realizing that the store brand is the cheapest in each category, does this exercise inform us about "store-brand-prone" households or "budget-prone" households? In other words, the fact that we observe households that purchase store brands across categories does not necessarily mean that there exist households with strong store-brand preferences—it could just mean that there exist cross-category price-sensitive households who buy the cheapest (i.e., store) brand. What is needed is a methodology that allows us to elicit the basic latent tendency for a household to buy the store brand across categories while controlling for marketing-mix variables. Making this distinction could be important for the retailer. For instance, if store-brand purchases are being made mainly due to store-brand preferences and not due to high price sensitivities, then the retailer could potentially increase category profits by narrowing the price gap between the store brand and the national brands. On the other hand, if households buy the store brand primarily because it is the cheapest brand, the retailer should carefully monitor the price gap between the store brand and national labels—especially in categories with competing "budget brands."

In this paper we develop a multicategory brand-choice model to study the extent to which there are "store-brand-prone" consumers that exhibit similarity in their preference for store brands across product categories. Using data from a large variety of product categories, we decompose household- and category-specific brand preferences and marketing-mix sensitivities into three components: one due to observed (demographic and shopping-related) household variables, another due to unobserved household-specific factors that are uncorrelated across categories, and one due to unobserved household-specific factors that are correlated across categories. These observed and unobserved components together capture the dependence across product categories. This implies a factor analytic structure on the covariance matrix of household unobservable heterogeneity. Imposing a factor structure allows us to not only characterize preferences in a lower-dimensional space, but potentially interpret the factors as inherent household "traits." Furthermore, if these factors are important in describing covariation in preferences across product categories (our results show this is indeed the case), then household-specific factor estimates can be used for targeting purposes and to predict demand for store brands in new product categories.

In this paper, we attempt to address the following questions:

- Are household preferences for store brands and other marketing-mix variables correlated across product categories? Does preference for store brand differ across broad groups of categories such as food versus nonfood?
- To what extent are these preferences driven by observed household characteristics versus unobserved household "traits?" How useful are the two components in identifying the *store-brand segment* and in predicting demand for store brands in *new categories*?
- Are store-brand households more or less price sensitive? Are the store brands for the retailer in question better positioned in certain categories?

Answers to these questions could be important to both retailers and the manufacturers of national brands. An understanding of the fundamental drivers of consumers' buying behavior can be a source of major strategic advantage. By developing a broader perspective of consumers' preferences and needs that transcends individual categories, managers can use their marketing resources more efficiently and effectively. For instance, finding strong correlations for

[1] This question of identifying the demographic or socioeconomic characteristics of store-brand buyers has long been of interest to marketing researchers (Frank and Boyd 1965, Rao 1969, Burger and Schott 1972, Szymanski and Busch 1987, Hoch 1996, Richardson et al. 1996). The findings from these studies have been mixed. Hoch (1996) uses store-level data and finds that store areas with higher percentages of elderly people; lower-income, larger families; and higher education have higher private-label shares. However, most other studies find that demographic characteristics of consumers do not explain the propensities to buy private labels.

[2] In our sample, the average correlation of the fraction of private-label purchases at the household level across 10 product categories is 0.22.

preference for store brands across categories indicates opportunities for the retailer to enhance their umbrella brand by leveraging the spillover effects of advertising and positioning from categories where their store brands command a strong position. Previous studies on umbrella branding suggest that consumers do transfer quality perceptions across products where the same brand name is used (Aaker 1991, Erdem 1998). Similarly, if significant correlations exist in preference for store brand across categories, then the parameter estimates could be useful in new-product targeting. Consider, for instance, a retailer who introduces a new private-label brand in a category. While the retailer may not have any information on the potential targets for this category, it can use information on households in other categories in which the retailer already sells the private label to infer potential targets for its brand in the new category. Targeting opportunities exist for the manufacturers of the national brands as well. For example, if private-label buyers are very price sensitive, then a very small dollar-value coupon could induce them to switch to their brand. Note that such targeting may not necessarily be in conflict with the retailer's interest. Depending on the product category, a switch to national brand could actually increase profits. Furthermore, the price-sensitive segment could also be most vulnerable to other competitors. Thus, by devising strategies such as manufacturer-sponsored in-store coupons, the retailer could potentially not only increase profits, but also enhance store loyalty.

The model is estimated using data from product categories that include a wide range of products such as canned food and condiments, as well as nonfood products like sandwich wraps and paper products. The data come from a frequent-shopper database from a large retail chain on the East Coast. Frequent-shopper data are unique in that they can capture the purchase behavior for a large number of customers and sometimes capture the entire customer base. As discussed in §3, the database we use captures over 90% of the store sales from the shopper card in most of the categories. This feature of the data makes this exercise particularly relevant from a managerial perspective. To illustrate, consider the problem of targeting a new store brand as discussed above. Suppose data are available only for a sample of customers. In this case, our task is not only to identify the target segments in the sample, but to also map those segments on some actionable demographic or lifestyle characteristics, which then becomes the basis of identifying potentials in the population. However, if the demographic variables turn out to be poor predictors of preferences (which is often the case; see Gupta and Chintagunta 1994, Rossi et al. 1996, and also in our application below), it impairs accessibility of segments, a key criterion for a good targeting and segmentation scheme (Wedel and Kamakura 1998). On the other hand, if purchase histories are available for the entire customer base, then the potential targets can be identified just using the preference parameters. Consistent with Rossi et al. (1996), we find that household-specific factor scores (estimated from the purchase history) are much more important in explaining covariation in preferences across product categories (and hence targeting) than the demographic variables.

There are also several disadvantages to the shopper card data used in this study. First, unlike the typical scanner panel data, we observe very little demographic information on the households. Due to privacy concerns, most retailers refrain from asking detailed demographic information at the time of card membership sign-up. A second disadvantage of frequent-shopper data is that households' purchases are observed only at the store in question. Thus, if a card holder in our sample shops at other stores, those purchases are not recorded. While this may seem like a major shortcoming, one must realize that this is the information that is typically available to the retailer (unless they purchase data from outside vendors like IRI or ACNielsen). Furthermore, in the context of the current study, purchase behavior from a single chain may in fact be preferred as the quality of the private label may vary significantly across chains. Thus, studies of private-label shopping behavior using typical scanner panel data could lead to findings that are severely biased due to contamination of the "brand-name" aspect of the private label.

We find significant correlation in households' preference for store brands and their price sensitivity across categories. All correlations are positive and significant with (nonweighted) average correlation of 0.32 for store brands, and 0.37 for price sensitivity. Interestingly, the average correlation for store-brand preference for nonfood categories is much higher (average of 0.47), indicating that households are more likely to exhibit a consistently strong preference for store brand across nonfood categories as compared to food categories. We also find that unobserved household factors are more important in explaining the variation in household preferences than are demographics. Using a two-dimensional factor structure, we find that the factors explain (on average) 42% for total variation in price sensitivity, and 34% of variation in store-brand preference.[3] More interestingly, one of the factors is found to explain a substantial amount of variation in store-brand preferences

[3] Corresponding numbers for eight demographics used in the study are 17% for price sensitivity and 14% for store-brand preference.

while the other factor explains price sensitivity—*consistently across categories*. The presence of these factors in all categories indicates that there are unobservable household-level components which are non-category specific, i.e., stable across categories. This in turn facilitates the interpretation of these components as unobservable household "traits."

Our applications show that, unlike demographics, household estimates of these factor scores can be very useful in market segmentation. For instance, we find the household estimates for the "price-sensitivity" factor to be very useful in identifying the most price-sensitive segment in the market. Similarly, we show that the "store-brand" factor is useful in not only identifying store-brand customers in the existing categories, but also in identifying targets for store brand in *new categories*. We demonstrate this using data from five holdout categories. Finally, our results show that store-brand buyers are more price sensitive. However, for the retailer in question, store brands are better positioned in certain categories as they attract buyers, mainly due to stronger preference rather than price reasons.

The rest of the paper is organized as follows. In §2 we describe our modeling approach and outline the main aspects of the estimation procedure. Section 3 describes the data used in the study. We present our empirical findings in §4. Section 5 demonstrates potential applications, and we conclude in §6.

2. Modeling Store-Brand Choice Across Categories

In this section we develop a multicategory brand-choice model to analyze the cross-category similarities in household preference for the store brands. Our main focus is on the estimation of the joint distribution of preferences and in estimating the degree of correlation in preferences across categories. The model proposed below allows us to understand the primary drivers of these correlations by decomposing household heterogeneity into category-specific components, observed household variables, and unobserved household-specific "factors."

Assume that each category is composed of three brands: a store brand, and two national brands.[4] For product category c, specify the utility for household h from purchasing the store brand, national brand 1, national brand 2, respectively, at purchase occasion t as

$$\begin{aligned} U_{hct,s} &= \beta_{hc,s} + \beta^p_{hc} p_{ct,s} + \beta^d_{hc} d_{ct,s} + \varepsilon_{hct,s}, \\ U_{hct,n_1} &= \beta_{hc,n_1} + \beta^p_{hc} p_{ct,n_1} + \beta^d_{hc} d_{ct,n_1} + \varepsilon_{hct,n_1}, \\ U_{hct,n_2} &= \beta_{hc,n_2} + \beta^p_{hc} p_{ct,n_2} + \beta^d_{hc} d_{ct,n_2} + \varepsilon_{hct,n_2}. \end{aligned} \tag{1}$$

Here s denotes store brand while n_1 and n_2 are national brands. The marketing-mix variables are price p and a dummy d indicating the presence of a promotion. We assume that the error terms in (1) follow an extreme value distribution leading to the well-known multinomial logit model.

For each category, one of the national brands is constrained to be the base brand.[5] Let β_{hc} denote the stacked vector of (identified) parameters for household h in category c (so β_{hc} is a four-dimensional vector in this example). A simple approach to modeling correlations in preferences across categories would be to estimate separate models for each category, and compute correlations in household estimates for the parameters (for example, this is the approach used in Kim et al. 1999). However, as Ainslie and Rossi (1998) point out, this two-step approach is not efficient and will underestimate the true cross-category correlations. Our approach is to let β_{hc} vary across households within categories due to their observed characteristics Z_h (i.e., household demographic information[6]) and unobserved heterogeneity u_{hc}:

$$\beta_{hc} = \Pi_c Z_h + u_{hc}, \quad c = 1, \ldots, C;\ h = 1, \ldots, H. \tag{2}$$

In the specification above, if the four-dimensional u_{hc}-vector was uncorrelated across categories, any correlations in the β_{hc}-vector across categories would have to be due to the demographic variables Z_h. This would rule out any stable unobservable preference effects across categories. As mentioned in the introduction, this does not seem realistic. At the same time, estimating a completely unrestricted covariance matrix cannot in general be recommended because the number of parameters quickly explodes. For instance, a 10-category model with four parameters each (this is the dimension in our empirical application) would result in 820 parameters in the covariance matrix itself. Our approach in the paper is to impose a factor structure and assume that the unobservable component u_{hc} varies due to (unobservable) category-specific effects u^*_{hc} and (unobservable) household specific effects ψ_h:

$$u_{hc} = \Gamma_c \psi_h + \Lambda_c u^*_{hc}, \quad c = 1, \ldots, C;\ h = 1, \ldots, H, \tag{3}$$

[4] The model can easily incorporate a varying number of brands across categories. Also note that we do not allow for a "no-purchase" option. While a more elaborate framework would model incidence for each category along with brand choice, this would add a substantial computational burden. We leave this open for future work.

[5] Specifically, we include a dummy for store brand and a dummy for the premium brand (defined as the highest-price brand) in each category.

[6] Note that Z_h also includes a constant.

where Λ_c is a diagonal matrix. Assuming that the u^*_{hc}-vector is uncorrelated across categories, this model allows for correlations in the β_{hc} across categories to arise from observed variables Z_h and unobserved variables ψ_h. If there are F elements in the ψ_h vector, their effect for category c is captured through the $4 \times F$ dimensional category-specific matrix Γ_c. Note that we allow the effects of both Z_h and ψ_h to be category specific. To see why this is important, suppose ψ_h is one-dimensional (i.e., $F = 1$) and is a measure of a household's "value consciousness." Then we could expect—for example—that ψ_h was important in explaining variation in both price sensitivities and store-brand preference in several categories and the corresponding elements of the Γ_c vector would be significant. However, it is conceivable that for some categories, households may suppress this element when choosing brands (for example, in certain food categories where nutritional or freshness attributes are more important than value). In this case, the relevant elements of Γ_c would be small.

For the specification (3) we assume a multivariate normal distribution for both ψ_h and u^*_{hc}: $\psi_h \sim \mathrm{N}(0, I_F)$ and $u^*_{hc} \sim \mathrm{N}(0, I_4)$. Together (2) and (3) model variation in β_{hc} as arising from three components Z_h, ψ_h, and u^*_{hc}, the effect of which measured through the matrices Π_c, Γ_c, and Λ_c. Two of these components (Z_h and ψ_h) are common across categories, whereas u^*_{hc} is category specific. Estimates of the parameters contained in $(\Pi_c, \Gamma_c, \Lambda_c)$ allow us to carefully measure which of the three components is most important in driving marketing-mix sensitivities and brand preferences in each of the categories. For example, one could compute the covariance induced by the factors as:

$$\mathrm{Cov}(\beta_h \mid z_h, \theta) = \begin{bmatrix} \Gamma_1\Gamma_1' + \Lambda_1\Lambda_1' & & & \\ \Gamma_2\Gamma_1' & \Gamma_2\Gamma_2' + \Lambda_2\Lambda_2' & & \\ \vdots & & \ddots & \\ \Gamma_C\Gamma_1' & \Gamma_C\Gamma_2' & \cdots & \Gamma_C\Gamma_C' + \Lambda_C\Lambda_C' \end{bmatrix}, \quad (4)$$

and to see how much of the correlation is driven by demographic information versus the factors, we can compute the unconditional covariance matrix:

$$\mathrm{Cov}(\beta_h \mid \theta) = \begin{bmatrix} \Pi_1\Omega_z\Pi_1' + \Gamma_1\Gamma_1' + \Lambda_1\Lambda_1' & & & \\ \Pi_2\Omega_z\Pi_1' + \Gamma_2\Gamma_1' & \Pi_2\Omega_z\Pi_2' + \Gamma_2\Gamma_2' + \Lambda_2\Lambda_2' & & \\ \vdots & & \ddots & \\ \Pi_C\Omega_z\Pi_1' + \Gamma_C\Gamma_1' & \Pi_C\Omega_z\Pi_2' + \Gamma_C\Gamma_2' & \cdots & \Pi_C\Omega_z\Pi_C' + \Gamma_C\Gamma_C' + \Lambda_C\Lambda_C' \end{bmatrix}, \quad (5)$$

where Ω_z is the covariance matrix of the demographic data: $\mathrm{Cov}(z_h) = \Omega_z$.

Note that a number of researchers have used a similar factor structure in logit (Elrod 1988, Chintagunta 1994, Erdem and Keane 1996, Erdem 1998) as well as probit (Elrod and Keane 1995, Keane 1997, Haaijer et al. 1998) settings. The primary motivation of imposing a factor structure in the majority of these papers is to study relative competition in a product category by producing a two-dimensional "brand map." In the context of the current application, an advantage of imposing a factor structure (besides parsimony) is that the factors can potentially be interpreted as inherent household "traits." For example, if we find the private-label parameters across many categories to load consistently on a certain factor, we can potentially label that as a "store-brand" factor. Furthermore, if these factors are important in describing covariation in preferences across product categories (our results show this is indeed the case), then household-specific factor estimates can be used for targeting purposes and to predict demand for store brands in new product categories.

The model described above allows us to measure the extent to which store-brand preference and marketing-mix sensitivities are correlated across categories and—if so—which components are driving these correlations. However, one should note that this is a nonlinear factor structure model (since the dependent variable, brand choice, depends nonlinearly on the factors). Choosing the number of factors and subsequently interpreting the factors can be quite challenging in nonlinear factor models. Below we report results for a two-factor specification, i.e., $F = 2$. We compare different measures of fit for models with one, two, and three factors and ultimately decide on a two-factor model (more on this in §4). Interpretation of the factors are then based on the estimated factor-loading matrices, Γ_c, $c = 1, \ldots, C$, and variance decompositions. Another feature of factor analytic structures is the indeterminacy of the rotation of the factor vector. We consider two alternative rotations: one that assumes that the store-brand and price-sensitivity factors are orthogonal, and another that is derived based on maximizing the store-brand factor and price-sensitivity factor's correlation with store-brand intercepts and price sensitivities, respectively. This second rotation leads to a nonorthogonal factor vector.

In the empirical application, we leave the loading matrices completely unrestricted except for the minimum restrictions required for identifying a unique factor rotation. In particular, when the number of factors is higher than one, we impose a triangularity restriction on a loading matrix for one category, i.e., on Γ_c, for some $c = 1, \ldots, C$. With two factors, this means one zero restriction on one of the elements of the second column of the chosen Γ_c. With three factors, this means three zero restrictions on the chosen Γ_c (so that a 3×3 submatrix of Γ_c is lower triangular). This still leaves the problem of *reflection*, i.e., that all loadings and factors can be multiplied by minus one without changing the posterior distribution. In theory, the implication of this indeterminacy are bimodal posteriors for loadings and factor scores. An easy solution to this problem is to impose sign restrictions on the loadings (e.g., impose a log-normal prior on one loading for each factor). However, for our application, without imposing any sign restrictions, we did not observe any switching between positive and negative loading signs once the Markov chain Monte Carlo (MCMC) sampler had converged to its stationary distribution. The Markov chain quickly settles around one mode and stays there. The reason for this is our very large sample. All parameters, including the factor loadings, have fairly tight posterior distributions. This precludes the sampler from switching the loading signs. For other data sets, where there is less information in the data, the sign restriction solution would be necessary to avoid bimodal posteriors. A similar problem, known as "label switching," arises in the estimation of mixtures; see Robert and Casella (1999). For further discussion, readers are referred to Anderson and Rubin (1956), Ansari et al. (2002), and Wedel and Kamakura (2001), who provide a useful discussion on the identification restrictions for a general class of such factor analytic models.

For inference, we use a hierarchical Bayesian approach. In particular, we use an MCMC procedure to simulate the posterior distribution of the model parameters and to compute household-level estimates of preferences. As discussed in Allenby and Rossi (1999), Bayesian procedures are well suited for these models, especially when one is interested in making inferences at the individual level. Because these procedures have become quite standard in the literature, we outline the estimation algorithm in the appendix. Interested readers are referred to (for example) Gelfand and Smith (1990) for a general overview of these methods, McCulloch and Rossi (1994) and Rossi et al. (1996) for multinomial probit models, and Allenby and Lenk (1994) for models where the error term follows an extreme value distribution. Estimation of multicategory models such as the one described above is an extension of the basic sampler laid out in these papers.

3. Data

The data used in the study comes from a frequent-shopper database from a large retail chain on the East Coast. We use data from a single store that is located in a small sub-urban town (confidentiality prevents us from revealing the name of the chain or town). Data are available for a period of 20 months from November 1999 to June 2001. We observe all transactions made in the store, including information such as time and date of the transaction, cardholder information (if shopper card is used), as well as dollar volume, unit price, quantity, and coupon usage for every UPC sold. The penetration of the shopper-card program is quite high for this store, with card holders accounting for over 83% of the total store sales.[7] Although noncard purchases account for 34% of all transactions, the average order size is significantly lower (average of $8 compared to $35 for transactions using a shopper card). In general, the noncard purchases tend to come from the coffee/snack shop and pharmacy; the shopper-card penetration rate on most grocery categories is well over 90% (including the categories used in this study).

Summary statistics for the product categories used in the analysis are reported in Table 1. These categories were selected to represent a wide range of products that would typically appear in a basket, and include canned and frozen food, condiments, other nonfood products such as sandwich wraps and cleaning supplies.[8] To keep our model and estimation manageable, we focus on the top two national brands and the private-label from each category.[9] Further, we aggregate over size and flavor for each brand. While this procedure may create some aggregation bias, we felt this to be necessary to include a wide range of product categories. However, in aggregating products to the brand level, we were careful not to include products for which the per-unit price varied significantly from other UPC's for that brand. In particular, we avoided certain very large or small size UPCs that are often sold at a significant discount or premium. For the remaining UPCs within a brand we ran correlation of prices to make sure the prices tended to move together over time. Finally, prices for the "brand" were created by share weighting the price of selected UPCs by its market share over the entire time period. The selected brands capture between 60% (bath) to well over 90% (oat) of the total category volume. We included all households that made at least one purchase from each of the 10 categories. This resulted in a total of 1,021 households with a total of 65,603 purchases.

[7] About 2% of these sales are on the employee card and thus cannot be traced back to any individual cardholder.

[8] Obviously, a more representative basket includes products from the produce and meat departments. However, the majority of these products are unbranded or carry the store-brand name with little other choice.

[9] While there is some evidence that retailers try to position their store's brands by imitating the top national brands (Sayman et al. 2002), it is concievable that store brands may be competing more closely with the brand closest to it in terms of price gap. At the suggestion of an anonymous reviewer, we conducted the entire analysis by including the highest-share national brand, national brand that was closest to the store brand in terms of price gap, and the store brand. The results were very similar and are available from the authors upon request.

Table 1 Summary Statistics for the 10 Categories Used for Estimation

Category	Brand	Share (%)	Price	Prom (%)	Category	Brand	Share (%)	Price	Prom (%)
Bath	Charmin	28.3	0.44	16.1	Oat	Cow	9.0	2.17	10.3
	Scott	35.2	0.62	10.3		Quaker	50.4	1.62	4.6
	Store	36.5	0.39	3.4		Store	40.6	1.18	8.0
Total number of observations = 10,512					Total number of observations = 4,957				
Foil	Reynolds	38.1	2.66	2.3	Paper	Bounty	42.5	1.91	24.1
	Saran	16.0	1.78	1.1		Brawny	21.3	1.55	14.9
	Store	45.9	1.21	10.3		Store	36.2	1.23	5.7
Total number of observations = 3,152					Total number of observations = 9,985				
Bacon	OM	29.1	4.56	9.2	Peanut	Jif	42.2	2.29	8.0
	Armour	26.4	3.10	11.5		Skippy	33.0	2.41	17.2
	Store	44.4	2.84	8.0		Store	24.8	1.75	1.1
Total number of observations = 6,189					Total number of observations = 4,985				
Mayo	Kraft	22.5	1.64	13.8	Waffle	AJ	30.4	1.71	13.8
	Hellmans	61.4	1.54	6.9		Eggo	50.5	1.97	10.3
	Store	16.1	0.95	1.1		Store	19.1	1.41	5.7
Total number of observations = 7,303					Total number of observations = 5,336				
Dish	Palmolive	45.1	2.18	6.9	Tuna	Star	34.6	0.95	9.1
	Dawn	41.7	2.22	17.2		BB	45.0	0.94	8.0
	Store	13.2	1.24	4.6		Store	20.4	0.56	5.7
Total number of observations = 5,497					Total number of observations = 7,687				

Consistent with the previous literature (Hoch and Banerji 1993, Raju et al. 1995), we find significant differences in store-brand penetration across categories. We also find that the private label is the lowest-priced brand in all categories. The promotion variables indicate the percentage of time the products were on a frequent-shopper discount. With minor exceptions, we find that the national brands were promoted more often than the store brand, perhaps to generate store traffic (Dreze 1995).

Table 2 presents the household-specific demographics and purchase-related variables used in the study. As discussed in the introduction, our data do not contain any demographic information. However, we do observe the mailing addresses for the card members. We use these addresses to obtain block-level demographic information. In addition, we use the household purchase history to create a number of purchase-related variables. The description of the household-specific variables is as follows: bask (average basket size of the card member), morning (% of shopping trips between 9 and 5 on weekdays), weekday (% of total trips on weekdays), income (average income), baby (dummy for presence of baby), pet (dummy for pet), milk (monthly milk expenditure—a proxy for household size), and HMR (monthly expenditure on home meal replacement—a measure of lifestyle).

Table 2 Summary Statistics for Demographic Information, 1,021 Households

	Mean	Std	Min	Max
Basket	65	38	12	391
Morning	0.38	0.21	0.00	0.92
Weekday	0.67	0.18	0	1
Income	51,016	22,449	5,000	145,000
Baby	0.17	—	—	—
Pet	0.36	—	—	—
Milk	9.6	8.8	0	87.0
HMR	3.8	5.6	0	50.5

With the exception of income, all other variables were created using the purchase history of the cardholders. For instance, if a household is observed to purchase diapers or baby food, it indicates the presence of an infant in the family. Similarly, purchases of dog food or cat litter indicate the presence of a pet. Note that several of these variables are proxies, as we do not observe the purchases of these households at other stores. We also do not hypothesize the relative importance or direction of these variables in explaining household preferences. Our primary motivation for including this wide range of demographic and purchase-related variables is to determine the extent to which they can be helpful in characterizing the store-brand buyers. We leave this as an empirical question.

4. Results

We now present the results from the model and data described above. We begin by describing the model fit and parameter estimates. This is followed by a discussion on the induced correlations in household preferences for store brands and marketing-mix sensitivities across categories.

Table 3 Model Fit

	In-sample fit hit rates and marginal likelihood			Out-of-sample fit hit rates		
	1 Factor	2 Factor	3 Factor	1 Factor	2 Factor	3 Factor
Bath	0.84	0.83	0.84	0.70	0.71	0.70
Dish	0.83	0.83	0.82	0.65	0.65	0.65
Foil	0.75	0.75	0.75	0.68	0.68	0.67
Bacon	0.75	0.75	0.75	0.62	0.62	0.62
Mayo	0.83	0.83	0.83	0.71	0.71	0.71
Oat	0.76	0.76	0.76	0.59	0.60	0.60
Paper	0.78	0.80	0.79	0.62	0.63	0.64
Peanut	0.82	0.83	0.83	0.68	0.68	0.68
Tuna	0.71	0.71	0.71	0.56	0.57	0.56
Waffle	0.74	0.74	0.73	0.61	0.61	0.61
Marginal log-likelihood	−27,807	−27,254	−27,231			

4.1. Model Fit and Estimates

We estimated one-, two-, and three-factor versions of the model described above. Table 3 shows in-sample and out-of-sample hit rates for all categories, along with estimates of the marginal log-likelihood for each of the three models (Newton and Raftery 1994). Based on the hit rates it is hard to discriminate between the three competing models: The fit is roughly the same. Parsimony would then dictate choosing the one-factor model. However, the marginal log-likelihood shows a huge increase in going from one factor to two. Going to a three-factor model, the marginal likelihood increases slightly, but by much less than the increase in going from one to two factors. Inspection of the estimates for the three-factor model (not shown, but available from the authors upon request) shows that the third factor only shows up significantly in a few categories. Based on this, we report estimates for the two-factor model below.

Tables 4 and 5 show estimates (posterior means and standard deviations) of the hierarchical coefficients contained in specifications (2) and (3). For ease of exposition and to conserve space, we present the results by attributes rather than by category. Table 4 shows the price-sensitivity estimates for all the categories. The mean price sensitivities (shown in the first column) have the expected negative sign in all categories (note that all demographic variables, apart from the two dummies "pet" and "baby," have been standardized to have mean zero and variance one). No consistent pattern is seen from the coefficient estimates to the demographic variables, and most of these are not significant in explaining variation in price sensitivities.[10] This is not surprising, as a number of previous studies have found demographic variables to be poor predictors of preferences (e.g., Rossi et al. 1996).

A more interesting story emerges from the estimated factor loadings. In particular, we find that the second factor loads strongly on all price equations with a negative sign (large positive factor scores imply high price sensitivity).[11] The first factor plays a much smaller role in explaining variation in price sensitivity. In the last four columns in Table 4, we decompose the overall variation in price sensitivity into variation attributed to the demographic variables ("Z"), the two factors (ψ_1, ψ_2), and residual unobserved heterogeneity (u). The variance decompositions are computed as follows. From (2) and (3), the price sensitivity in category c, $\beta_{hc,p}$, is modeled as

$$\beta_{hc,p} = \pi_{c,p} Z_h + \gamma_{c,p,1}\psi_{h1} + \gamma_{c,p,2}\psi_{h2} + \lambda_{c,p} u^*_{hc,p},$$

where $\pi_{c,p}$ and $(\gamma_{c,p,1}, \gamma_{c,p,2})$ are the rows of Π_c and Γ_c corresponding to price, and $\lambda_{c,p}$ is the diagonal element of Λ_c corresponding to price. This implies a variance equal to

$$\mathrm{V}(\beta_{hc,p}) = \pi_{c,p}\Omega_Z\pi'_{c,p} + \gamma^2_{c,p,1} + \gamma^2_{c,p,2} + \lambda^2_{c,p},$$

where Ω_Z is the covariance matrix for the demographic variables. This decomposes overall variance into components due to demographics, the two factors, and residual heterogeneity. For example, the fraction explained by demographic information is

$$r_{c,p,Z}(\theta) \equiv \frac{\pi_{c,p}\Omega_Z\pi'_{c,p}}{\pi_{c,p}\Omega_Z\pi'_{c,p} + \gamma^2_{c,p,1} + \gamma^2_{c,p,2} + \lambda^2_{c,p}}.$$

The table shows the posterior mean of $r_{c,p,Z}(\theta)$ (and similarly for the fraction explained by the two factors and residual heterogeneity). In spite of including a large number of household observed variables (total of eight "Z" variables), they explain only (on average) 17% of the overall variation. Similarly, the first factor contributes very little to variation in price sensitivities (from 1% to 10%). The second factor is found to be much more important: This factor explains from 21% (for the "bath" category) to 61% (for the "peanut" category) of overall variation.

The parameter estimates for the store-brand intercept are presented in Table 5. Recall from the discussion in the data section that we use three brands in each category. In our specification, we include two dummies: store brand and one national brand (the highest-priced brand for the category, which we refer to as the premium brand). Note that the store-brand

[10] In the tables, we have represented "significant" estimates in bold, where by "significant" we mean that either at least 97.5% of the posterior mass is concentrated on the positive axis or 97.5% is concentrated on the negative real axis.

[11] As always in factor analysis, the loadings and factor scores are only determined up to a sign. In other words, we can always multiply loadings and factor scores with minus one so that all the price loadings are positive.

Table 4 Posterior Estimates of Hierarchial Coefficients for Price Attribute

		Demographics (Z)								Factors (ψ)		u	Variance decomposition			
Category	Const.	Bsize	PMT	PWT	INC	Baby	Pet	Milk	HMR	γ_1	γ_2	λ	Z	ψ_1	ψ_2	u
Bath	**−5.01**	−0.55	−0.76	−0.90	−0.61	0.14	−0.02	0.59	−0.09	−1.31	**−2.87**	**4.86**	0.14	0.06	0.21	0.59
	(0.92)	(0.62)	(0.86)	(0.91)	(0.56)	(1.31)	(1.15)	(0.55)	(0.58)	(0.76)	(0.63)	(0.84)	(0.06)	(0.05)	(0.08)	(0.11)
Dish	**−3.04**	−0.25	0.11	**−0.76**	0.05	−0.72	0.34	−0.16	−0.31	0.02	**−1.51**	**1.53**	0.18	0.02	0.39	0.41
	(0.39)	(0.25)	(0.34)	(0.34)	(0.23)	(0.62)	(0.49)	(0.24)	(0.24)	(0.30)	(0.31)	(0.28)	(0.07)	(0.02)	(0.11)	(0.11)
Foil	**−1.22**	0.37	−0.05	0.09	−0.37	0.06	0.30	0.28	−0.30	−0.32	**−1.15**	**0.92**	0.30	0.07	0.37	0.25
	(0.45)	(0.29)	(0.39)	(0.39)	(0.27)	(0.79)	(0.59)	(0.29)	(0.30)	(0.42)	(0.39)	(0.23)	(0.13)	(0.09)	(0.16)	(0.11)
Bacon	**−1.83**	0.10	−0.26	0.19	0.11	0.09	0.20	−0.03	0.08	−0.11	**−0.98**	**0.93**	0.07	0.01	0.48	0.43
	(0.15)	(0.09)	(0.13)	(0.13)	(0.09)	(0.23)	(0.18)	(0.09)	(0.09)	(0.13)	(0.12)	(0.09)	(0.03)	(0.02)	(0.08)	(0.08)
Mayo	**−3.11**	0.20	0.33	−0.31	0.05	0.99	0.09	**0.85**	0.09	−0.39	**−2.03**	**2.22**	0.19	0.03	0.35	0.43
	(0.60)	(0.39)	(0.57)	(0.54)	(0.37)	(0.90)	(0.71)	(0.40)	(0.40)	(0.48)	(0.49)	(0.25)	(0.07)	(0.04)	(0.11)	(0.10)
Oats	**−2.81**	−0.02	0.31	**−0.77**	0.20	0.96	**1.24**	**−0.82**	0.05	−0.10	**−1.30**	**2.05**	0.23	0.02	0.22	0.54
	(0.39)	(0.26)	(0.35)	(0.34)	(0.25)	(0.64)	(0.50)	(0.25)	(0.25)	(0.34)	(0.34)	(0.18)	(0.06)	(0.02)	(0.09)	(0.09)
Paper	**−1.73**	−0.26	0.07	−0.25	0.56	0.45	−0.39	**0.77**	0.17	**0.91**	**−1.50**	**2.18**	0.18	0.10	0.23	0.49
	(0.49)	(0.32)	(0.44)	(0.45)	(0.30)	(0.72)	(0.59)	(0.33)	(0.31)	(0.41)	(0.44)	(0.20)	(0.07)	(0.07)	(0.10)	(0.09)
Peanut	**−4.27**	0.66	−0.19	−0.30	−0.21	−0.27	−0.48	0.49	0.27	0.35	**−2.16**	**0.96**	0.23	0.03	0.61	0.14
	(0.46)	(0.28)	(0.43)	(0.40)	(0.26)	(0.67)	(0.54)	(0.28)	(0.27)	(0.37)	(0.40)	(0.35)	(0.08)	(0.04)	(0.11)	(0.08)
Tuna	**−1.93**	0.08	−0.13	0.40	0.03	−0.56	−0.73	0.14	0.05	0.04	**−2.60**	**3.11**	0.04	0.01	0.39	0.56
	(0.37)	(0.24)	(0.34)	(0.34)	(0.23)	(0.60)	(0.46)	(0.24)	(0.25)	(0.34)	(0.33)	(0.24)	(0.02)	(0.01)	(0.07)	(0.07)
Waffle	**−1.96**	−0.31	0.50	−0.50	−0.22	0.65	−0.26	−0.13	0.26	−0.41	**−1.81**	**1.28**	0.13	0.04	0.54	0.28
	(0.37)	(0.23)	(0.34)	(0.31)	(0.23)	(0.56)	(0.44)	(0.24)	(0.24)	(0.31)	(0.30)	(0.31)	(0.06)	(0.05)	(0.11)	(0.11)

Note. "Significant" model parameters (defined as parameters whose posterior distribution has at least 97.5% of the posterior mass to one side of zero) are set in boldface type.

name is common to all categories, while the other brand constant will be unique to each category (for example, Kraft in mayonnaise and Bounty in paper towel). For the store-brand intercept, we find that the first factor loads strongly with a negative sign for all categories. Given our normalization, a large negative first-factor score implies a strong preference for the store brand. This first factor explains from 11%

Table 5 Posterior Estimates of Hierarchial Coefficients for Private-Label Attribute

		Demographics (Z)								Factors (ψ)		u	Variance decomposition			
Category	Const.	Bsize	PMT	PWT	INC	Baby	Pet	Milk	HMR	γ_1	γ_2	λ	Z	ψ_1	ψ_2	u
Bath	−0.27	**−0.35**	−0.26	−0.15	**−0.21**	0.14	0.26	**0.19**	0.01	**−1.28**	0.00	**1.40**	0.10	0.41	0.00	0.49
	(0.16)	(0.10)	(0.14)	(0.15)	(0.10)	(0.26)	(0.19)	(0.09)	(0.10)	(0.12)	(na)	(0.09)	(0.03)	(0.05)	(na)	(0.05)
Dish	**−4.42**	**−0.62**	−0.08	−0.26	−0.23	−0.62	0.32	0.01	−0.05	**−1.74**	**−0.72**	**1.66**	0.11	0.42	0.08	0.39
	(0.37)	(0.21)	(0.28)	(0.29)	(0.18)	(0.49)	(0.39)	(0.19)	(0.20)	(0.27)	(0.29)	(0.17)	(0.04)	(0.08)	(0.06)	(0.07)
Foil	−0.12	−0.03	−0.19	0.16	**−0.41**	0.19	0.29	0.24	−0.24	**−1.44**	−0.28	**1.32**	0.14	0.44	0.03	0.38
	(0.32)	(0.20)	(0.27)	(0.28)	(0.19)	(0.55)	(0.42)	(0.20)	(0.21)	(0.29)	(0.29)	(0.15)	(0.06)	(0.10)	(0.04)	(0.08)
Bacon	**−0.24**	−0.03	−0.18	**0.21**	−0.07	0.00	0.14	0.02	0.01	**−0.38**	**−0.22**	**0.67**	0.09	0.20	0.08	0.63
	(0.10)	(0.06)	(0.09)	(0.09)	(0.06)	(0.16)	(0.12)	(0.06)	(0.06)	(0.08)	(0.09)	(0.05)	(0.04)	(0.07)	(0.05)	(0.08)
Mayo	**−2.43**	−0.13	−0.11	0.00	−0.22	0.52	−0.13	**0.52**	0.02	**−1.27**	0.13	**1.35**	0.17	0.38	0.02	0.44
	(0.31)	(0.19)	(0.29)	(0.27)	(0.19)	(0.45)	(0.35)	(0.20)	(0.20)	(0.22)	(0.26)	(0.12)	(0.06)	(0.09)	(0.03)	(0.08)
Oats	**−0.81**	−0.03	0.13	−0.29	−0.03	0.31	**0.54**	**−0.40**	0.09	**−0.63**	0.07	**0.77**	0.25	0.29	0.02	0.43
	(0.18)	(0.12)	(0.16)	(0.17)	(0.12)	(0.30)	(0.24)	(0.12)	(0.12)	(0.16)	(0.17)	(0.08)	(0.08)	(0.11)	(0.03)	(0.09)
Paper	**−0.50**	**−0.29**	−0.12	−0.11	0.02	0.49	0.05	**0.46**	−0.04	**−0.85**	−0.08	**0.88**	0.23	0.36	0.02	0.39
	(0.19)	(0.13)	(0.17)	(0.17)	(0.11)	(0.28)	(0.23)	(0.13)	(0.12)	(0.16)	(0.18)	(0.08)	(0.06)	(0.09)	(0.02)	(0.08)
Peanut	**−1.78**	0.13	−0.19	0.13	−0.25	0.04	−0.41	0.12	0.18	**−0.92**	−0.37	**1.40**	0.09	0.26	0.05	0.59
	(0.23)	(0.14)	(0.22)	(0.21)	(0.13)	(0.35)	(0.28)	(0.15)	(0.14)	(0.18)	(0.21)	(0.10)	(0.04)	(0.07)	(0.05)	(0.08)
Tuna	**−1.37**	**−0.33**	−0.14	0.14	−0.01	−0.41	−0.16	0.06	−0.09	**−0.41**	−0.17	**1.06**	0.13	0.11	0.03	0.73
	(0.12)	(0.08)	(0.11)	(0.11)	(0.07)	(0.20)	(0.15)	(0.08)	(0.08)	(0.10)	(0.11)	(0.07)	(0.04)	(0.05)	(0.03)	(0.06)
Waffle	**−1.25**	−0.17	0.03	−0.09	−0.17	0.19	−0.21	−0.06	0.05	**−0.54**	−0.19	**1.09**	0.09	0.18	0.03	0.70
	(0.15)	(0.10)	(0.14)	(0.13)	(0.09)	(0.23)	(0.18)	(0.09)	(0.09)	(0.11)	(0.13)	(0.08)	(0.03)	(0.06)	(0.03)	(0.07)

Note. "Significant" model parameters (defined as parameters whose posterior distribution has at least 97.5% of the posterior mass to one side of zero) are set in boldface type.

Table 6 Posterior Estimate of Correlation Matrix for Price Sensitivities

Attribute	Bath	Dish	Foil	Bacon	Mayo	Oats	Paper	Peanut	Tuna	Waffle
Bath										
Dish	**0.33** (0.09)									
Foil	**0.33** (0.12)	**0.36** (0.13)								
Bacon	**0.33** (0.08)	**0.42** (0.08)	**0.43** (0.12)							
Mayo	**0.32** (0.09)	**0.34** (0.10)	**0.43** (0.13)	**0.43** (0.08)						
Oats	**0.22** (0.09)	**0.36** (0.09)	0.24 (0.13)	**0.35** (0.08)	**0.24** (0.10)					
Paper	0.19 (0.10)	**0.29** (0.10)	0.22 (0.14)	**0.32** (0.09)	**0.32** (0.11)	0.17 (0.10)				
Peanut	**0.39** (0.11)	**0.49** (0.10)	**0.51** (0.14)	**0.55** (0.08)	**0.53** (0.10)	**0.33** (0.11)	**0.46** (0.11)			
Tuna	**0.27** (0.07)	**0.37** (0.07)	**0.37** (0.10)	**0.43** (0.06)	**0.36** (0.08)	**0.25** (0.08)	**0.30** (0.08)	**0.49** (0.07)		
Waffle	**0.39** (0.10)	**0.47** (0.10)	**0.44** (0.14)	**0.50** (0.08)	**0.46** (0.11)	**0.37** (0.09)	**0.30** (0.11)	**0.54** (0.10)	**0.46** (0.07)	

Note. "Significant" model parameters (defined as parameters whose posterior distribution has at least 97.5% of the posterior mass to one side of zero) are set in boldface type.

to 44% of the overall variation in store-brand preferences. The second factor plays a much smaller role in explaining store-brand preferences. It should also be noted that category-specific components are quite important, reflecting, perhaps, category-specific differences such as store-brand quality. Finally, we find that apart from the "oats" and "paper" categories, demographics do not explain much in terms of brand preferences.

While we do not report the parameter estimates for promotion and premium brand to conserve space (they are available from the authors), we provide a brief discussion on the main findings. For promotion sensitivities all intercepts were found to have the expected positive sign, with the largest effects of promotions in the "tuna" and "bacon" categories. Similar to the price and private-label parameters, most of the demographic coefficients were small. The first factor had relatively small loadings for promotion (except for the "paper" category), while all signs for the second-factor loadings were found to be positive (or zero). Thus, large positive scores for the second factor imply high price sensitivity and high sensitivity to promotions. As for the premium-brand preference, we found Factor 1 to enter with a consistent positive sign (this is the opposite sign Factor 1 had in the store-brand equations). Factor 2 enters with a positive or zero loading. However, neither of these factors explained a substantial amount of the overall variation for the national brands in any category. Instead, a majority of the variation is found to be category specific, which seems reasonable because none of the national brand names are common across these 10 categories.

4.2. Correlations in Preferences

The presence of the common factors (along with demographics) leads to correlations in sensitivities and preferences across categories, which are reported in Tables 6 and 7. The reported correlations are computed in the following way. Consider, for example, the correlation in price sensitivities between two categories, c and c', i.e., $\mathrm{Cor}(\beta_{hc,p}, \beta_{hc',p})$. From (2) and (3),

$$\beta_{hc,p} = \pi_{c,p} Z_h + \gamma_{c,p}\psi_h + \lambda_{c,p} u^*_{hc,p},$$
$$\beta_{hc',p} = \pi_{c',p} Z_h + \gamma_{c',p}\psi_h + \lambda_{c',p} u^*_{hc',p}.$$

The implied covariance matrix for $(\beta_{hc,p}, \beta_{hc',p})$ is then

$$\mathrm{V}(\beta_{hc,p}, \beta_{hc',p} \mid \theta) = \begin{pmatrix} \pi_{c,p}\Omega_Z\pi'_{c,p}+\gamma_{c,p}\gamma'_{c,p}+\lambda^2_{c,p} & \\ \pi_{c,p}\Omega_Z\pi'_{c',p}+\gamma_{c,p}\gamma'_{c',p} & \pi_{c',p}\Omega_Z\pi'_{c',p}+\gamma_{c',p}\gamma'_{c',p}+\lambda^2_{c',p} \end{pmatrix}.$$

Then

$$\begin{aligned}\rho_{c,c',p}(\theta) &\equiv \mathrm{Cor}(\beta_{hc,p}, \beta_{hc',p} \mid \theta)\\ &= (\pi_{c,p}\Omega_Z\pi'_{c',p} + \gamma_{c,p}\gamma'_{c',p})\\ &\quad \cdot \big((\pi_{c,p}\Omega_Z\pi'_{c,p} + \gamma_{c,p}\gamma'_{c,p} + \lambda^2_{c,p})^{1/2}\\ &\quad \cdot (\pi_{c',p}\Omega_Z\pi'_{c',p} + \gamma_{c',p}\gamma'_{c',p} + \lambda^2_{c',p})^{1/2}\big)^{-1}.\end{aligned}$$

Using the MCMC draws of θ, we can then simulate the implied posterior distribution of $\rho_{c,c',p}(\theta)$. The tables

Table 7 **Posterior Estimate of Correlation Matrix for Store-Brand Attribute Preferences**

Attribute	Bath	Dish	Foil	Bacon	Mayo	Oats	Paper	Peanut	Tuna	Waffle
Bath										
Dish	**0.47** (0.05)									
Foil	**0.47** (0.06)	**0.49** (0.08)								
Bacon	**0.31** (0.06)	**0.38** (0.07)	**0.37** (0.07)							
Mayo	**0.46** (0.05)	**0.40** (0.08)	**0.46** (0.08)	**0.28** (0.07)						
Oats	**0.35** (0.08)	**0.35** (0.10)	**0.32** (0.11)	**0.21** (0.08)	**0.26** (0.10)					
Paper	**0.48** (0.06)	**0.43** (0.08)	**0.46** (0.09)	**0.30** (0.07)	**0.49** (0.07)	**0.24** (0.10)				
Peanut	**0.35** (0.06)	**0.38** (0.07)	**0.39** (0.07)	**0.30** (0.06)	**0.35** (0.07)	**0.22** (0.09)	**0.34** (0.07)			
Tuna	**0.24** (0.05)	**0.31** (0.06)	**0.25** (0.07)	**0.20** (0.05)	**0.20** (0.07)	**0.12** (0.07)	**0.22** (0.06)	**0.19** (0.06)		
Waffle	**0.30** (0.06)	**0.35** (0.06)	**0.32** (0.07)	**0.23** (0.06)	**0.27** (0.07)	**0.23** (0.08)	**0.27** (0.07)	**0.26** (0.06)	**0.19** (0.05)	

Note. "Significant" model parameters (defined as parameters whose posterior distribution has at least 97.5% of the posterior mass to one side of zero) are set in boldface type.

show the posterior mean and standard deviation of the correlations. We only discuss correlations for the price and private-label parameters. We also found positive (although small) correlations in the promotion parameter, but the table is not reported here to conserve space.

Table 6 shows the correlations of price sensitivities across categories. All price sensitivity correlations are positive and of a substantial size. Several correlations are above 0.50. The (nonweighted) average correlation is 0.37. Computing these correlations with the household estimates (not reported here), this average jumps to 0.58. In general, these correlations are much higher than those reported in the previous literature (for example, Ainslie and Rossi (1998) found an across-category price sensitivity correlation of 0.28).

Table 7 shows the across-category correlations for store-brand preference. As for price sensitivity, these correlations are all positive and quite large. The average correlation is 0.32 (this number jumps to 0.47 with household estimates). Only three correlations are less than 0.20. It is of interest to note that the correlations between the categories "bath," "dish," "foil," and "paper" are all greater than 0.40. The average correlation between these categories is 0.47. These four categories are the four nonfood categories included in the sample. This result indicates that households are more likely to exhibit a consistently strong preference for store brand across nonfood categories as compared to food categories. There could be several reasons for such a finding. For instance, it is possible that the quality variation in nonfood categories is smaller, especially if the store brand is supplied by the same manufacturer. Similarly, while we do not find any common brand names across the 10 categories used (except store brand), several of the national brands in the nonfood categories are by the same manufacturer (for example, Bounty, Charmin, and Dawn all belong to P&G). At the same time, food products are more likely to involve idiosyncratic tastes where households may find the quality/taste of store brand acceptable in some categories, but not in others.[12]

5. Applications

5.1. Interpreting the Factors

Discussion of the results above shows that imposing a factor structure can be quite useful in characterizing preferences in a lower-dimensional space. While our approach does not require that the factors be interpreted—they can be seen simply as a tool

[12] Another potential explanation for the finding is that we treat the purchase decisions in different categories as independent and ignore any possible substitutability or complementarities of consumption needs across categories. For instance, it is conceivable that certain food products such as bacon and tuna, or waffle and oat, are substitutes in consumption usage. In general, if food categories (compared to nonfood categories) are more substitutable in consumption usage, a household who is very price sensitive but also has a strong preference for national brands may switch between food categories depending on which one has promotions for national brands. By ignoring this possible substitutability we may be underestimating the correlations of household preferences among food categories. We thank an anonymous reviewer for pointing this out.

to model dependence—it is potentially relevant from a marketing perspective to label the factors. Based on the results presented above, we could label Factor 1 as the "store-brand" and Factor 2 the "price-sensitivity" factor. This interpretation is based on the factor loadings in Tables 4 and 5, and the resulting variance decompositions. The labeling should not be taken to represent that the first factor only affects store-brand preferences and the second only price sensitivities (in fact, both factors enter the price and store-brand equations). The naming of the factors simply refers to the finding above that the first factor explains a substantial amount of variation in store-brand preferences, while the second factor explains substantial variation in price sensitivity—*consistently across categories*.

This labeling of the factors seems reasonable under the restriction that the store-brand and price factors are orthogonal (as imposed in our estimation). However, given that our aim is to interpret the store-brand and price factor as unobservable household "traits," it seems somewhat arbitrary to restrict the factors to be orthogonal. It is conceivable that households' inherent preference for store brand may be correlated with their inherent price sensitivity. Note also that both factors enter the store-brand and price sensitivity equations significantly for many categories. We next consider an alternative approach based on a rotation of the original factors, which maximizes the interpretation of the factors.[13]

Consider first the store-brand intercepts across all categories. From (2) and (3) we have:

$$\beta_{h,sb} = \Pi_{sb} Z_h + \Gamma_{sb}\psi_h + \Lambda_{sb} u^*_{h,sb}, \tag{6}$$

where $\beta_{h,sb}$ is the stacked store-brand intercepts across all categories for household h, and Π_{sb}, Γ_{sb}, Λ_{sb} are composed of the relevant elements of the Π_c, Γ_c, and Λ_c matrices for $c = 1, \ldots, 10$. Our interest centers on the 10×2 loading matrix Γ_{sb} and the factor vector ψ_h. One way to get the "ideal" store-brand factor is to rotate the original factor vector ψ_h to get a new store-brand factor (say $\psi^*_{h,sb} = \rho'_{sb}\psi_h$) so as to have the *maximum correlation* with the $\beta_{h,sb}$ vector. In other words, we seek $\psi^*_{h,sb}$ and an associated loading vector γ^*_{sb} such that

$$\Gamma_{sb}\psi_h \approx \gamma^*_{sb}\psi^*_{h,sb}. \tag{7}$$

This is really a problem of finding the best "Rank-1" approximation to $\Gamma_{sb}\psi_h$.[14] To obtain the best Rank-1 approximation, we may use the singular value decomposition of Γ_{sb}, given by

$$\Gamma_{sb} = UDV', \tag{8}$$

where U is a 10×2 orthogonal matrix containing the left singular vectors, D is a 2×2 diagonal matrix containing the singular values, and V is a 2×2 orthogonal matrix containing the right singular vectors. The first row of V' contains the linear combination of the factor vector, which maximizes correlation with $\Gamma_{sb}\psi_h$. We chose ρ'_{sb} as this first row. Using a similar argument for the price sensitivities, we obtain a second rotation $\rho'_p\psi_h$ (using a second singular-value decomposition) which maximizes correlation with the price sensitivity parameters. So the new rotation is

$$\psi^*_h \equiv \rho\psi_h = \begin{pmatrix} \rho'_{sb}\psi_h \\ \rho'_p\psi_h \end{pmatrix}. \tag{9}$$

The implied covariance matrix of the rotated (nonorthogonal) factors is

$$\mathrm{V}[\psi^*_h] = \rho\rho'. \tag{10}$$

Because by definition the ρ-vectors have unit length, this matrix will have 1s along the diagonal. Note that ρ is a function of the loading matrices Γ_c, $c = 1, \ldots, 10$. So we can easily get posterior draws of ρ and a mean posterior estimate of ρ (and of course posterior distributions of the household-specific rotated factor scores). Figure 1 shows 50 draws of the rotation vectors and the posterior mean rotation (calculated using all draws from the sampling algorithm). We see that the optimal rotation is in fact nonorthogonal. The implied correlation between the two optimal factors is 0.31 (the off-diagonal element in (10)). The signs of the rotated factor loadings (not shown) imply that households with large values of the (rotated) price factor are very price sensitive, and households with large values of the (rotated) store brand factor have the smallest store-brand intercepts.[15] We next discuss how household estimates of the rotated factor scores can be used for making marketing decisions.

5.2. Using Factor Scores for Market Segmentation

The procedure described above provides us with household-level estimates for two factors: one that captures household's inherent price sensitivity and the second that captures the store-brand preference. These household-level estimates can be quite useful from a practical perspective. For instance, suppose the retailer is interested in segmenting customers based on their price sensitivity or store-brand preference. Previous research in marketing (and our results above) has shown that demographics in general are not very useful for that purpose. On the other hand,

[13] We are indebted to the area editor and an anonymous referee for suggesting this.

[14] Note that $\Gamma_{sb}\psi_h$ is the product of a 10×2 and a two-dimensional vector, while $\gamma^*_{sb}\psi^*_{h,sb}$ is the product of a 10×1 matrix and a scalar.

[15] These results indicate that the correlation between store-brand intercepts and price sensitivities is (small but) positive. However, store-brand buyers are still found to be more price sensitive than national brand buyers; see Table 8.

Figure 1 Fifty Draws of the Rotation Matrix and the Mean Posterior Rotation

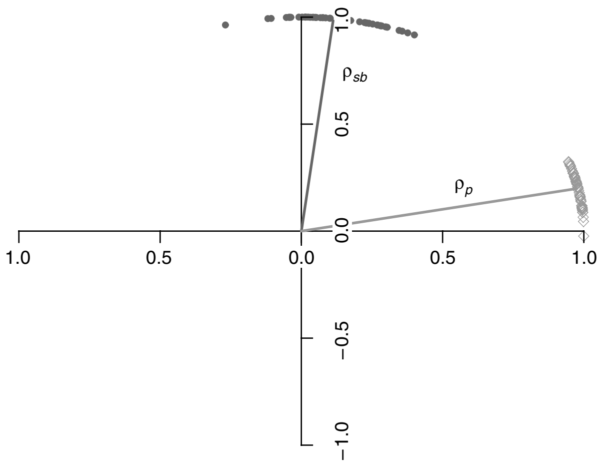
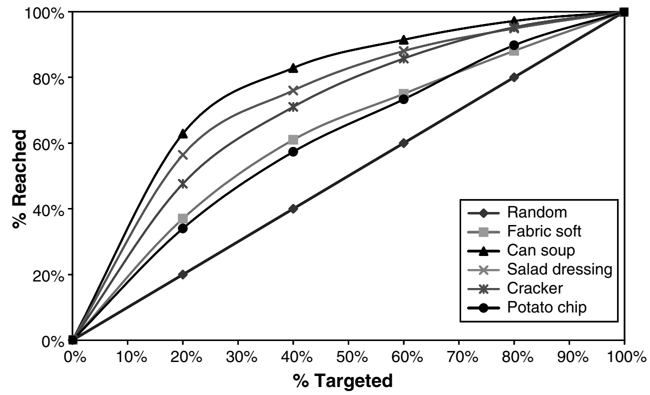


Figure 2 Targeting Store-Brand Buyers in Holdout Categories



preference estimates from household purchase history have been shown to be very useful (for example, see Rossi et al. 1996 for a single-category case). To test whether estimates of the factor scores discussed above can be useful in segmenting customers across a broad range of product categories, we assigned the households in our sample into quartiles based on the empirical distribution of the factor scores. In all 10 categories in our sample we found the factor scores to do a very good job in rank ordering customers (these numbers are not reported, but are available from the authors). In particular, the average price sensitivity for households in the top quartile based on the price sensitivity factor was found to be at least three times that of households in the bottom quartile (in every category). Similarly, store-brand shares for households in the top quartile based on the store-brand factor were found to be at least twice that of households in the bottom quartile (shares were 5 to 10 times higher in many categories).

Another important question for the retailer is whether household estimates for the store-brand factor can be useful in identifying targets for store brand in a *new category*. For instance, suppose the retailer in question introduces a private label in a new category. Can the household estimates for the store-brand factor be used to infer potential targets in this new category? While we do not observe new private-label introduction in our data, we demonstrate the potential application using five holdout categories. In each category a household was assigned as a potential target if it had the highest share for store brand over the entire purchase history. Our objective would be to test the extent to which a single estimate of the store-brand factor can be useful in targeting these individuals. In Figure 2 we show the resulting gains charts. The 45-degree line represents the performance if targeting is done at random. Gains due to information from the store-brand factor are reflected in the extent to which the line lies above the 45-degree line. The chart shows that targeting 20% of the population based on the highest values of store-brand factor would include 37% (fabric softener), 63% (canned soup), 56% (salad dressing), 48% (cracker), and 34% (potato chips) of our target audience. By contrast, 20% of the population picked at random would include only 20% of the target group. Thus, we find that household estimates for the factor scores cannot only be useful in identifying the store-brand segment in the existing products, but also in targeting store brands in new product categories.

5.3. Drivers of Store-Brand Choice Across Categories

So far we have focused on assessing the similarities in household purchase behavior across categories. We now turn to the question of drivers of store-brand choice in individual categories. In particular, we address the following question: Are store-brand purchases being driven by brand preferences or due to price considerations? In other words, if we consider the segment consisting of households who have high probability of store-brand purchase, do households then belong to this segment primarily because of a high store-brand intercept (relative to national label intercepts) or because of the price component or both? Note that the answer to this may vary with category: In categories where some households perceive the store brand to be of high quality, price considerations may be a less important driver than in categories where the store brand is perceived to be of poor quality. This question is of key importance to the retailer. If store-brand purchases are being driven by the price component only to a small degree, then the retailer can narrow the price gap between the store-brand and the national brands. Similarly, categories that attract store-brand buyers mainly due to stronger preferences could be used as flagship categories for carrying the store "brand name."

Table 8 Drivers of Store-Brand Choice

	Price sensitivity		Decomposition of SB share (%)		
Category	NB	SB	Total	SB-Loyal	SB-Price
Bath	−4.92	−5.44	31	80	20
Dish	−2.88	−4.01	9	20	80
Foil	−0.80	−1.33	49	56	44
Bacon	−1.11	−2.08	53	34	66
Mayo	−2.54	−5.01	16	47	53
Oat	−1.18	−3.35	37	66	34
Paper	−0.85	−3.31	42	56	44
Peanut BT	−4.32	−5.59	24	33	67
Tuna	−1.74	−5.70	18	14	86
Waffle	−1.89	−3.03	11	19	81

To answer this question, we classify households in each category as either store-brand households (defined as households whose maximum choice probability is for store brand) or national-brand households (defined as the complement of the store-brand households). The first question we ask is whether the store-brand households are more, or less, price sensitive? Note that households whose maximum purchase probability is for store brand will tend to have higher store-brand intercepts and/or higher price sensitivities than national-brand households.[16] In Table 8 we report the mean price sensitivity for the store- and national-brand households as defined above. In every category, we find that store-brand households are more price sensitive than their national-brand counterparts. While researchers have long speculated this to be the case (Hoch 1996, Dhar and Hoch 1997, Kalyanam and Putler 1997, Boatwright et al. 2004), this has not been empirically documented.

In the last two columns of Table 8, we further divide the store-brand households into two segments: SB-Loyal and SB-Price defined as follows:

$$\begin{aligned} \mathrm{SB}_{\mathrm{Loyal}} &= 1 \quad \text{if } \{\beta_{\mathrm{SB}}\} > \max\{\beta_{\mathrm{NB}j}\},\ j=1,2 \\ \mathrm{SB}_{\mathrm{Price}} &= 1-\mathrm{SB}_{\mathrm{Loyal}}. \end{aligned} \tag{11}$$

In other words, we assign households to be store-brand loyal if their store-brand intercept is the highest in the category. If not, we classify them as SB-Price, i.e., they choose store brand because of price reasons. As seen from Table 8, there are significant differences across categories. The store brand seems to best positioned in the bath tissue category (and to a certain extent in oat), as it attracts 80% of its buyers due to stronger preference. On the other hand, store brand is most vulnerable in dish, tuna, and waffle categories, as the majority of the households buy the store brand due to price reasons.

6. Conclusion

Store brands form an important element of a retailer's marketing-mix strategy. The focus of this paper is on understanding the behavior of store-brand buyers. Using unique frequent-shopper data, we develop a multicategory brand-choice model. The vector of household-level preferences for attributes and sensitivities to marketing variables is modeled as a function of observable household characteristics (demographic variables) and a small number of unobservable household-specific "factors." These observed and unobserved household-specific components together capture the dependence across product categories. Results from 10 product categories indicate that households do display similarities in their preference for store brands and marketing-mix sensitivity across categories. Furthermore, the proposed factor structure is quite useful in eliciting the basic latent tendency for a household to buy the store brand across categories. Household-level factor estimates could be used in identifying the "store-brand-prone" consumers and targeting these households for store brands in new product categories.

There are, of course, several caveats to our study. Foremost, our study uses data from a single store that raises concerns about the generality of the results. We also do not observe household purchases at competing stores, which is problematic because recent research (Shankar and Bolton 2004) has shown competition to be the most important factor in retailer pricing decisions, and hence household purchase behavior. Second, to keep the estimation manageable, we focussed on only the top three brands in each category and aggregated UPCs to a brand level that could bias the estimates. Similarly, while our finding that demographics are not useful in identifying the store-brand buyers is consistent with the previous literature, most of the demographic variables were created using the purchase-history data or using census block data. Finally, our model is a conditional choice model, and we do not capture a household's purchase incidence decisions. This could be an important issue particularly if one is interested in extending the current analysis to applications such as optimal pricing and profitability analysis. While theoretically a no-purchase option can easily be incorporated in the current framework, it requires additional assumptions

[16] To realize this is straightforward: The set of brand intercepts and price sensitivities implying that store brand has the highest choice probability is

$$R_s = \{\beta\colon \beta_s - \beta_{nb_1} > \beta^p(p_{nb_1} - p_s),\ \beta_s - \beta_{nb_2} > \beta^p(p_{nb_2} - p_s)\}$$

or

$$R_s = \left\{\beta\colon \beta^p < \min\left\{\frac{\beta_s - \beta_{nb_1}}{p_{nb_1} - p_s}, \frac{\beta_s - \beta_{nb_2}}{p_{nb_2} - p_s}\right\}\right\}.$$

The distribution of β truncated to the set R_s will have a smaller mean β^p and larger mean β_s than the nontruncated distribution.

about the purchase cycle for each category and poses significant estimation challenges in a multicategory setting. However, with better data and advancements in computing powers, these shortcomings could be overcome in the future.

This study also points out several directions for future work. Our results suggest that store brands are better positioned in certain categories, as they attract buyers primarily due to strong preferences rather than price concerns. In these categories, retailers could consider narrowing the price gap with national brands. Previous research on determining the optimal price gap between private-label and national brands could be extended to multicategory settings, especially in cases of complimentary categories (Manchanda et al. 1999). Similarly, our results suggest that household estimates could be used in predicting demand for store brands in new categories. While we test this in holdout categories, it would be useful to do such an exercise where there is an actual product launch. Given the high failure rates of new products in general, such targeting could be quite useful in the initial phases of the product launch.

Acknowledgments
The authors thank the editor, area editor, and four anonymous reviewers for their valuable comments and suggestions.

References

Aaker, David. 1991. Managing brand equity: Capitalizing on the value of a brand name. Free Press, New York.

Ainslie, A., P. Rossi. 1998. Similarities in choice behavior across multiple categories. *Marketing Sci.* **17**(2) 91–106.

Allenby, G., P. Lenk. 1994. Modeling household purchase behavior with logistic normal regression. *J. Amer. Statist. Assoc.* **89**(December) 1218–1231.

Allenby, G., P. Rossi. 1991. Quality perceptions and asymmetric switching between brands. *Marketing Sci.* **10**(3) 185–204.

Allenby, G. M., P. E. Rossi. 1999. Marketing models of consumer heterogeneity. *J. Econometrics* **89** 57–78.

Anderson, T. W., H. Rubin. 1956. Statistical inference in factor analysis. J. Neymann, ed. *Proc. Third Berkeley Sympos. Math. Statist. Probab.*, Vol. 5. University of California Press, Berkeley, CA, 111–150.

Ansari, A., K. Jedidi, L. Dube. 2002. Heterogeneous factor analysis models: A Bayesian approach. *Psychometrika* **67**(1) 49–78.

Blattberg, R., K. Wisniewski. 1989. Price-induced patterns of competition. *Marketing Sci.* **8**(4) 291–309.

Boatwright, P., S. Dhar, P. Rossi. 2004. The role of retail competition, demographics, and account retail strategy as drivers of promotions sensitivity. *Quant. Marketing Econom.* **2** 169–190.

Burger, P., B. Schott. 1972. Can private brand buyers be identified? *J. Marketing Res.* **9**(2) 219–222.

Chintagunta, P. 1994. Heterogeneous logit model implications for brand positioning. *J. Marketing Res.* **31** 304–311.

Chintagunta, Pradeep, Andre Bonfrer, Inseong Song. 2002. Investigating the effects of store-brand introduction on retailer demand and pricing behavior. *Management Sci.* **48**(10) 1242–1267.

Corstjens, Marcel, Rajiv Lal. 2000. Building store loyalty through store brands. *J. Marketing Res.* **37** 281–291.

Cotterill, R., W. Putsis, R. Dhar. 2000. Assessing the competitive interaction between private labels and national brands. *J. Bus.* **73** 109–137.

Dhar, Sanjay, Stephen Hoch. 1997. Why store brand penetration varies by retailers. *Marketing Sci.* **3** 208–227.

Dreze, X.. 1995. Loss leaders: Store traffic and cherry picking. Unpublished Ph.D. thesis, University of Chicago, Chicago, IL.

Elrod, T. 1988. Choice map: Inferring a product market map from panel data. *Marketing Sci.* **7** 21–40.

Elrod, T., M. P. Keane. 1995. A factor-analytic probit model for representing the market structure in panel data. *J. Marketing Res.* **32** 1–16.

Erdem, T. 1998. An empirical analysis of umbrella branding. *J. Marketing Res.* **35**(3) 339–351.

Erdem, Tulin, Michael Keane. 1996. Decision-making under uncertainty: Capturing dynamic brand choice processes in turbulent consumer goods markets. *Marketing Sci.* **15**(1) 1.

Frank, R., H. Boyd. 1965. Are private-brand-prone grocery consumers really different? *J. Advertising Res.* **5**(4) 27–35.

Gelfand, A., A. Smith. 1990. Sampling-based approaches to calculating marginal densities. *J. Amer. Statist. Assoc.* **85** 398–409.

Gupta, S., P. Chintagunta. 1994. On using demographic variables to determine segment membership in logit mixture models. *J. Marketing Res.* **31**(1) 128–136.

Haaijer, Rinus, Michel Wedel, Marco Vriens, Tom Wansbeek. 1998. Utility covariances and context effects in conjoint MNP models. *Marketing Sci.* **17**(3) 236–252.

Hoch, Stephen. 1996. How should national brands think about private labels? *Sloan Management Rev.* **37**(2) 89–102.

Hoch, Stephen, Shumeet Banerji. 1993. When do private labels succeed? *Sloan Management Rev.* **34**(4) 57–68.

Kalyanam, Kirthi, Daniel Putler. 1997. Incorporating demographic variables in brand choice models: An indivisible alternatives approach. *Marketing Sci.* **16**(2) 166–181.

Keane, Michael. 1997. Modeling heterogeneity and state dependence in consumer choice behavior. *J. Bus. Econom. Statist.* **15**(3) 310.

Kim, B., K. Srinivasan, R. Wilcox. 1999. Identifying price sensitive consumers: The relative merits of demographic vs. purchase pattern information. *J. Retailing* **75**(2) 173–193.

Manchanda, P., A. Ansari, S. Gupta. 1999. The shopping basket: A model for multi-category purchase incidence decisions. *Marketing Sci.* **18** 95–114.

McCulloch, R., P. Rossi. 1994. An exact likelihood approach to analysis of MNP model. *J. Econometrics* **64** 207–240.

Mills, David. 1995. Why retailers sell private labels. *J. Econom. Management Strategy* **4**(3) 509–528.

Newton, M., A. Raftery. 1994. Approximate Bayesian inference with the weighted likelihood bootstrap. *J. Roy. Statist. Soc., Ser. B* **56** 3–48.

Pauwels, K., S. Srinivasan. 2004. Who benefits from store brand entry? *Marketing Sci.* **23**(3) 364–390.

Raju, Jagmohan, Raj Sethuraman, Sanjay Dhar. 1995. The introduction and performance of store brands. *Management Sci.* **41** 957–978.

Rao, T. R. 1969. Are some consumers more prone to purchase private brands? *J. Marketing Res.* **6**(4) 447–450.

Richardson, P., A. Jain, A. Dick. 1996. Household store brand proneness: A framework. *J. Retailing* **72**(2) 159–185.

Robert, C., G. Casella. 1999. Monte Carlo statistical methods. Springer-Verlag, New York.

Rossi, P., R. McCulloch, G. Allenby. 1996. On the value of household purchase history information in target marketing. *Marketing Sci.* **15**(4) 321–340.

Sayman, Serdar, Stephen Hoch, Jagmohan Raju. 2002. Positioning of store brands. *Marketing Sci.* **21**(4) 378–397.

Scott-Morton, Fiona, Florian Zettelmeyer. 2000. The strategic positioning of store brands in retailer-manufacturer bargaining. Working paper, Yale University, New Haven, CT.

Seetharaman, P. B., A. Ainslie, P. Chintagunta. 1999. Investigating household state dependence effects across categories. *J. Marketing Res.* **36** 488–500.

Sethuraman, Raj. 1992. The effect of marketplace factors on private label penetration in grocery products. Report #92, Marketing Science Institute, Boston, MA.

Shankar, V., R. Bolton. 2004. An empirical analysis of determinants of retailer pricing strategy. *Marketing Sci.* **23**(1) 28–49.

Singh, Vishal, Karsten Hansen, Sachin Gupta. 2005. Modeling preferences for common attributes in multi-category brand choice. *J. Marketing Res.* Forthcoming.

Szymanski, D., P. Busch. 1987. Identifying the generics-prone consumer: A meta-analysis. *J. Marketing Res.* **24**(4) 425–431.

Wedel, M., W. A. Kamakura. 1998. *Market Segmentation: Conceptual and Methodological Foundations*, *International Series in Quantitative Marketing*. Kluwer Academic Publishers, Dordrecht, The Netherlands.

Wedel, M., W. Kamakura. 2001. Factor analysis with mixed observed and latent variables in the exponential family. *Psychometrika* **66**(4) 515–530.