SIDDHARTHA CHIB, P.B. SEETHARAMAN, and ANDREI STRIJNEV*

In usual practice, researchers specify and estimate brand-choice models from purchase data, ignoring observations in which category incidence does not occur (i.e., no-purchase observations). This practice can be problematic if there are unobservable factors that affect the no-purchase and the brand-choice decisions. When such a correlation exists, it is important to model simultaneously the no-purchase and the brand-choice decisions. The authors propose a model suitable for scanner-panel data in which the no-purchase decision depends on the price, feature, and display of each brand in the category and on household stock of inventory. They link the no-purchase model to the brand-choice outcome through marketing-mix covariates and through unobservables that affect both outcomes. The authors assume that model parameters are heterogeneous across households and allow for a flexible correlation structure between the coefficients in the no-purchase model and those in the brand-choice model. The model formulation is more general than what is possible from either a nested logit model or a translog utility model and from models in which the no-purchase outcome is an additional outcome with the deterministic component of its utility set equal to zero. The authors estimate the proposed model using Bayesian Markov chain Monte Carlo estimation methods. They then apply the estimation methods to scanner-panel data on the cola product category and compare the results with those from the widely used nested logit model.

Model of Brand Choice with a No-Purchase Option Calibrated to Scanner-Panel Data

There is a vast literature in marketing that pertains to the estimation of brand-choice models (e.g., Bass 1974; Chintagunta, Jain, and Vilcassim 1991; Gönül and Srinivasan 1993; Guadagni and Little 1983; Massy, Montgomery, and Morrison 1970; Rossi and Allenby 1993). The attention devoted to these models is a testament both to the complexity of the brand-choice problem and to the value that the models have for understanding the impact of marketing-mix variables (e.g., price, feature, display) on individual choices. In the panel-data context, appropriately specified brand-choice models also provide an understanding of heterogeneity in brand-choice across households, which aids, for example, in the development of differentiated product lines and target-marketing initiatives.

However, the issue of the no-purchase outcome and its proper modeling has received only marginal attention in the literature. To set up this problem, note that almost all the published brand-choice formulations (whether in the multinomial logit or multinomial probit family) address only those occasions when a purchase occurs; occasions when no-purchase occurs (and consequently no brand-choice is observed) are dropped from the analysis. This widely adopted approach (and the only approach that textbooks discuss) is justifiable only if the unobservable factors that drive the no-purchase decision are not correlated with factors that influence brand choice, conditioned on the marketing-mix variables. If the unobservables are correlated, the data on the no-purchase occasions have value for estimating the brand-choice parameters, even though the brand-choice information is missing for the observations. Following usage in statistics, we refer to this case as one of informative-missingness (Little and Rubin 1987).

The purpose of this article is to develop a brand-choice model in which the no-purchase (binary) outcome is carefully modeled and allowed to correlate with the unobservables that influence brand choice. We cast the model in Bayesian terms and estimate it using Markov chain Monte

*Siddhartha Chib is Harry C. Hartkopf Professor of Econometrics and Statistics (e-mail: chib@wustl.edu), and P.B. Seetharaman is Assistant Professor of Marketing (seetharaman@olin.wustl.edu), John M. Olin School of Business, Washington University. Andrei Strijnev is Assistant Professor of Marketing, School of Management, University at Buffalo (e-mail: strijnev@buffalo.com). The authors are listed in alphabetical order.

Journal of Marketing Research
Vol. XLI (May 2004), 184–196

Carlo (MCMC) methods (Chib 2001; Chib and Greenberg 1995, 1996) based on the approach of Albert and Chib (1993). In our model, we specify the household's no-purchase decision to depend not only on the household's stock of inventory in the product category but also on the price, feature, and display of each brand in the category in a completely flexible manner (as we explain subsequently).

The next part of our model provides the link between the no-purchase decision and the brand-choice model. We specify the link directly through the common marketing-mix covariates and through common unobservables that affect both outcomes. Whether such common unobservables are present is an empirical question that can be answered only by a model specification that allows for such dependence and a comparison of that model with one that does not. Finally, we complete our model by specifying brand choice through the usual random-utility formulation. This portion of the model is standard; however, note that we assume that the coefficients in the brand-choice model are heterogeneous across households, as are those in the no-purchase model. In addition, we specify the heterogeneity distribution so that the coefficients in the no-purchase model can correlate in a completely general way with the coefficients in the brand-choice model. We believe that allowing for such correlation is important, because any restrictions on the correlations are difficult to support on a priori grounds.

Our model formulation is quite distinct and more general than existing approaches, which we summarize as follows: First, the nested logit model achieves the link between a household's brand-choice and category-purchase outcomes through a composite measure of product-category attractiveness (i.e., inclusive value measure). This measure restricts the relative importance of marketing variables (e.g., price versus feature) to remain the same between a household's brand-choice and no-purchase decisions. However, it is possible, for example, that a newspaper feature advertisement influences a household's decision to purchase a product category, but it may have no role in the household's brand choice within the category. The inclusive value measure also restricts the brand with the highest baseline preference in the brand-choice model to have the most influence on a household's no-purchase decision. However, a household can differentially focus on salient brands to decide whether to buy the product but then evaluate all brands equally in its brand-choice decision. The inclusive value measure also ignores unobserved correlations between a household's no-purchase and brand-choice outcomes that arise from common unobservables. For example, a household can systematically purchase private label brands for parties or when expecting guests for a weekend. Because such occasions are not observed in scanner-panel data, their effects can be captured only by a flexible correlation structure between the household's no-purchase and brand-choice utilities, which is not possible in the nested logit model. For these reasons, which we subsequently explain in detail, our proposed model is more general than the nested logit model.

Second, for reasons that are similar to those discussed in the previous paragraph, our model formulation is also more general than the translog utility models of Hanemann (1984), Chiang (1991), Chintagunta (1993), and Arora, Allenby, and Ginter (1998). Although the translog utility model allows for correlations between no-purchase and brand-choice outcomes that arise from unobservables in the household's translog utility function that simultaneously translate into the first-order conditions for the household's category-purchase and brand-choice decisions, it imposes a correlation structure that is in line with translog utility theory. Such a correlation structure is quite restrictive. However, our model allows unobserved correlations between the two decisions to be as general as possible. Furthermore, our model captures the effects of brand-specific covariates on the two household decisions in a flexible manner (as we explain in the previous paragraph).

Third, our model is also distinct from a specification in which the no-purchase outcome is modeled simply as an additional outcome. This approach has been used in a few articles (especially in the emerging literature on new empirical industrial organization; see, e.g., Chintagunta 2002) in which the deterministic component of utility attached to the no-purchase outcome is set equal to 0. Our proposed model dominates such a multinomial add-one-outcome approach, which is even more parametrically restrictive than the nested logit and translog utility models.

The rest of the article is organized as follows: In the next section, we present the proposed model and discuss some of its properties. We then discuss the application of the model to the analysis of scanner-panel data from Information Resources Inc. on the cola product category. Together with the results from the estimation, we show the empirical gains from our model in relation to those from the nested logit formulation. We also provide a discussion of managerial implications and the use of the model in brand pricing. In the Appendix, we develop our Bayesian fitting procedure.

MODEL OF BRAND CHOICE WITH NO-PURCHASE OPTION

To develop our model of brand choice with the no-purchase option, suppose that the typical household h (h = 1, 2, ..., H), observed over t = 1, 2, ..., n_h shopping occasions, either buys or does not buy one of J brands in a specific category. On any given shopping occasion, we observe two outcome variables y_{ht} and y^*_{ht}, where y_{ht} is a binary outcome variable that takes the value of 1 if a purchase in that category occurs on that shopping occasion and takes the value of 0 otherwise. When $y_{ht} = 1$ (i.e., category purchase occurs), y^*_{ht} takes the value j, j = 1, 2, ..., J. When $y_{ht} = 0$, the outcome y^*_{ht} is missing. As is typical in scanner-panel data sets, we assume that for every shopping visit t, we observe the price (P_{htj}), display (D_{htj}), and feature (F_{htj}) covariates that the household faces, regardless of whether it purchases in the category. The goal is to model the outcome variables (y_{ht}, y^*_{ht}) given the previous information.[1]

The modeling approach that we propose has two distinct components. The first models the binary outcome y_{ht}, and the second models the multinomial outcome y^*_{ht}. Common unobservable factors that affect both outcomes link the two components. More formally, let z_{ht} denote the (indirect) utility of household h at time t at the category level; we assume that we can marginally express this utility as a function of the entire set of category–brand covariates that the household faces and as its composite stock of category inventory (I_{ht}) as follows:

[1]We ignore the modeling of purchase quantity.

$$(1)\quad z_{ht} = \gamma_{1h} + \gamma_{2h}P_{ht1} + \gamma_{3h}P_{ht2} + ... + \gamma_{(J+1)h}P_{htJ} + \gamma_{(J+2)h}D_{ht1} + \gamma_{(J+3)h}D_{ht2} + ... + \gamma_{(2J+1)h}D_{htJ} + \gamma_{(2J+2)h}F_{ht1} + \gamma_{(2J+3)h}F_{ht2} + ... + \gamma_{(3J+1)h}F_{htJ} + \gamma_{(3J+2)h}I_{ht} + \nu_{ht},$$

where $\gamma_h = (\gamma_{1h}, \gamma_{2h}, ..., \gamma_{(J+1)h}, ..., \gamma_{(2J+2)h}, ..., \gamma_{(3J+1)h}, \gamma_{(3J+2)h})$ are household-specific coefficients (weights) of each factor, and ν_{ht} is a normally distributed random variable with mean of zero and variance of one. We determined the category-level outcome y_{ht} according to the sign of z_{ht} as follows:

$$(2)\quad y_{ht} = I[z_{ht} > 0],$$

where I[A] is the indicator function that takes the value of 1 when event A occurs and the value of 0 otherwise. In other words, the household buys into the category if the category utility is sufficiently high, crossing the threshold of zero. We would expect that, all else being equal, higher prices depress utility (and thus reduce the probability of category purchase), display and feature enhance utility (and thus increase the probability of category purchase), and an inventory increase reduces utility. Note that we allow the category utility to depend on each of the category–brand covariates, which is more general than using a share-weighted average of marketing variables across brands (e.g., Chib, Seetharaman, and Strijnev 2002; Manchanda, Ansari, and Gupta 1999) or an inclusive value term that is constructed from the estimates of a conditional brand-choice model (e.g., Bucklin and Gupta 1992).

Now, let u_{hjt} denote the (indirect) utility of household h for brand j and shopping visit t. We assume that we can marginally express this utility as a function of the entire set of brand-specific covariates that the household faces as follows:

$$(3)\quad u_{htj} = \alpha_{hj} + \delta_{2h}P_{htj} + \delta_{3h}D_{htj} + \delta_{4h}F_{htj} + \eta_{htj},$$

where α_{hj}, j = 1, 2, ..., J, are brand-specific and household-specific intercepts (for identification purposes, $\alpha_{h1} = 0$), and $\delta_h = (\alpha_{h2}, \alpha_{h3}, ..., \alpha_{hJ}, \delta_{2h}, \delta_{3h}, \delta_{4h})'$ are household-specific coefficients in the brand model. We assume that the errors $\eta_{ht} = (\eta_{ht1}, ..., \eta_{htJ})$ in this brand-choice model are normally distributed with mean of zero and the covariance matrix $\Sigma = \text{diag}(1, \sigma_{22}, ..., \sigma_{JJ})$.

We avoid introducing correlations between the brand utilities because correlations are difficult to identify when J is large, as in our subsequent examples. This is not appreciated in the multinomial probit literature, primarily because researchers have not attempted to fit high-dimensional multinomial probit models with fully unrestricted covariances. It is not difficult to determine that under our assumption, the first diagonal element of Σ is one, and the parameters $\sigma = (\sigma_{22}, ..., \sigma_{JJ})$ are identified.

We determine the brand-level outcome y^*_{ht} in the usual way: by the principle of maximum utility. We observe the outcome $y^*_{ht} = j$ when the utility of the jth brand exceeds that of the remaining brands. Specifically, $y^*_{ht} = j$ iff $u_{htj} > \max_{k \neq j} u_{htk}$.

We would expect that, all else being equal, higher prices depress utility (and thus reduce the household's probability of brand purchase) and that display and feature enhance utility (and thus increase the household's probability of brand purchase). Note that we assume that the price, display, and feature coefficients are constant across all brands in the product category. This is consistent with the specification of conditional brand-choice models in the literature.

For identification purposes, we further rewrite the household's brand-specific indirect utilities in difference form as

$$(4)\quad u^{\dagger}_{htj} = \alpha_{hj} + \delta_{2h}P^{\dagger}_{htj} + \delta_{3h}D^{\dagger}_{htj} + \delta_{4h}F^{\dagger}_{htj} + \eta^{\dagger}_{htj},$$

where $P^{\dagger}_{htj} = P_{htj} - P_{ht1}$, $D^{\dagger}_{htj} = D_{htj} - D_{ht1}$, and $F^{\dagger}_{htj} = F_{htj} - F_{ht1}$ are the covariates expressed in difference form, relative to Brand 1, and $\eta^{\dagger}_{htj} = \eta_{htj} - \eta_{ht1}$ is the differenced error term. It follows that the vector of differenced errors $\eta^{\dagger}_{ht} = (\eta^{\dagger}_{ht2}, ..., \eta^{\dagger}_{htJ})$ in this brand-choice model is normally distributed with mean of zero and the following covariance matrix:

$$(5)\quad \Sigma^{\dagger} = \begin{pmatrix} 1+\sigma_{22} & 1 & \cdots & 1 \\ 1 & 1+\sigma_{33} & \cdots & \vdots \\ \vdots & \ddots & \ddots & 1 \\ 1 & \cdots & 1 & 1+\sigma_{JJ} \end{pmatrix}.$$

Correspondingly, we observe the outcome $y^*_{ht} = j$ for j > 1 when the transformed utility of the jth brand exceeds not only that of the remaining brands but also zero. Specifically:

$$(6)\quad y^*_{ht} = j \text{ iff } u^{\dagger}_{htj} > 0 \text{ and } u^{\dagger}_{htj} > \max_{k \neq j} u^{\dagger}_{htk} \quad (j > 1),$$

and we observe the outcome $y^*_{ht} = 1$ when the differenced utilities of all brands j = 2, ... J are less than zero.

We combine the two components of the models by making the following two general and testable assumptions: The first assumption is that the category-specific errors ν_{ht} are correlated with the brand-specific transformed errors $\eta^{\dagger}_{ht}$. In particular, we assume that

$$(7)\quad \left(\nu_{ht}, \eta^{\dagger\prime}_{ht}\right)' \sim N_J(0, \Omega^{\dagger}),$$

where

$$(8)\quad \Omega^{\dagger} = \begin{pmatrix} 1 & \rho_{12} & \rho_{13} & \cdots & \rho_{1J} \\ \rho_{12} & 1+\sigma_{22} & 1 & \cdots & 1 \\ \rho_{13} & 1 & 1+\sigma_{33} & \cdots & \vdots \\ \vdots & \vdots & 1 & \ddots & \vdots \\ \rho_{1J} & 1 & \cdots & \cdots & 1+\sigma_{JJ} \end{pmatrix}$$

is the covariance matrix of the combined errors. Of particular interest in Ω are the parameters $\rho = (\rho_{12}, ..., \rho_{1J})'$, where $\rho_{1j} = \text{Cov}(\nu_{ht}, \eta^{\dagger}_{htj})$ is the covariance between the household's category-purchase and brand-choice decisions caused by common unobserved factors that affect both outcomes. This dependence has not been demonstrated in previous brand-choice models with a no-purchase outcome. Second, we assume that the household-specific coefficients in the no-purchase and brand-choice models are jointly normally distributed:

$$(9)\quad \beta_h = (\gamma_h, \delta_h)' \sim N_{4\times(J+1)}(\beta, D),$$

where the matrix D, which is of dimension 4(J + 1) × 4(J + 1), is completely unrestricted. If we partition D as

$$(10)\quad D = \begin{pmatrix} D_{11} & D_{12} \\ D'_{12} & D_{22} \end{pmatrix},$$

where $D_{11} = \text{Var}(\gamma_h)$ is of dimension $(3J + 2) \times (3J + 2)$ and $D_{22} = \text{Var}(\delta_h)$ is of dimension $(J + 2) \times (J + 2)$, a key matrix of interest is $D_{12} = \text{Cov}(\gamma_h, \delta_h)$, which is a matrix of dimension $(3J + 2) \times (J + 2)$. This matrix gives the covariances between the coefficient heterogeneity in the no-purchase model and the coefficient heterogeneity in the brand-choice model. For example, the elements in the D_{12} matrix that give the covariance between the price coefficients in the category-purchase and brand-choice specifications would help isolate whether a sensitivity to price in the no-purchase decision is also associated with a similar sensitivity to price in the brand-choice decision. Although some research has focused on whether a given household responds similarly to price across product categories (Ainslie and Rossi 1998; Seetharaman, Ainslie, and Chintagunta 1999), to our knowledge, no study has investigated this question in the context of the brand-choice model with a no-purchase outcome.

Previous Models

Our proposed brand-choice model with a no-purchase option provides a flexible characterization of households' choice behavior in a product category. We compare our model with some previously proposed ways of handling the no-purchase option:

The simplest way to accommodate the no-purchase option is to include an additional alternative in the J-alternative multinomial logit (MNL) model to obtain a (J + 1)–alternative MNL model, where the (J + 1)th alternative refers to a household's option not to purchase any brand in the product category (Chintagunta 2002). In this model, the (indirect) utilities of household h for the J brands and the no-purchase option during shopping visit t are represented, respectively, as follows:

(11) $$u_{htj} = \alpha_{hj} + \delta_{2h}P_{htj} + \delta_{3h}D_{htj} + \delta_{4h}F_{htj} + \eta_{htj}, \text{ and}$$

(12) $$u_{ht0} = \eta_{ht0},$$

where the errors $\eta_{ht} = (\eta_{ht1}, ..., \eta_{htJ}, \eta_{ht0})'$ are distributed i.i.d. Gumbel with a scale parameter of 1. In this model, we simultaneously determine the brand-level outcome y^*_{ht} and the category-level outcome y_{ht} by the principle of maximum utility (i.e., we observe the outcome $y^*_{ht} = j$ when the utility of the jth brand exceeds that of the remaining brands and that of the no-purchase option, and we observe the outcome $y_{ht} = 1$ when the utility of the no-purchase option exceeds that of all the brands).

This model suffers from two key limitations: First, it ignores the separate effects of covariates on the household's decision to purchase the category. Second, by assuming that the brand-specific and no-purchase-option errors are i.i.d., the model imposes a restrictive pattern of household switching between brands and the no-purchase option (especially because no-purchase outcomes vastly outnumber brand-choice outcomes in any scanner-panel data set).

Another way to accommodate the no-purchase option is to allow the household's random utility for the (J + 1)th alternative to depend on brand-specific covariates and to assume a flexible correlation structure of the errors, in addition to the inclusion of a (J + 1)th alternative in the MNL model. The nested logit model achieves this (Bucklin and Gupta 1992). In this model, the household's brand-choice probability is a product of two probabilities: the household's category-purchase probability and its conditional brand-choice probability. Furthermore, the household's category-purchase probability depends linearly on the sum of the household's exponentiated deterministic brand-specific utilities (also called the "inclusive value" of the product category). In this model, we express the (indirect) utility of household h at time t at the category level (z_{ht}) in the following way:

(13) $$z_{ht} = \gamma_{1h} + \gamma_{2h}IV_{ht} + \gamma_{3h}I_{ht} + \nu_{ht},$$

where $IV_{ht} = \ln\{\exp(z_{ht1}) + \exp(z_{ht2}) + ... + \exp(z_{htJ})\}$ is the inclusive value, which depends on the household's brand-specific deterministic utilities z_{htj}, which we express as follows:

(14) $$u_{htj} = z_{htj} + \eta_{htj} = \alpha_{hj} + \delta_{2h}P_{htj} + \delta_{3h}D_{htj} + \delta_{4h}F_{htj} + \eta_{htj},$$

where the brand-level errors $\eta_{ht} = (\eta_{ht1}, ..., \eta_{htJ})$ and the category-level error ν_{ht} are all distributed Gumbel with a scale parameter of 1. As in our proposed model, we determine the category-level outcome y_{ht} according to the sign of z_{ht}: $y_{ht} = I[z_{ht} > 0]$, where I[A] is the indicator function that takes the value of 1 when event A occurs and the value of 0 otherwise. In addition, as in our proposed model, we observe the brand-level outcome $y^*_{ht} = j$ when the utility of the jth brand exceeds that of the remaining brands. Specifically, $y^*_{ht} = j$ iff $u_{htj} > \max_{k \neq j} u_{htk}$.

This nested logit model is different from our proposed model in two important ways: First, the error from the category-purchase model ν_{ht} and the errors from the brand-choice model $\eta_{ht} = (\eta_{ht1}, ..., \eta_{htJ})$ are not correlated. Second, the household's indirect utility at the category level depends on the brand-specific marketing variables through the single-summary inclusive value measure. The former feature rules out unobserved dependencies between the household's category and brand-choice decisions, whereas the latter feature implies that—conditioned on the parameters δ_{2h}, δ_{3h}, and δ_{4h}—the marketing variables have identical effects on the household's no-purchase decision (through the parameter γ_{2h}), which is a restrictive assumption in general. Furthermore, the second feature also implies that any asymmetry across brands in terms of the effect on the household's purchase-incidence decision is captured only through the asymmetric brand-specific intercepts (parameters α_{hj}) in the brand-choice model. This means that the brand that has the highest (lowest) baseline preference for the household is necessarily the brand that most (least) influences the household's purchase-incidence decision in the product category.

An alternative to the nested logit model is a brand-choice model derived from economic primitives of direct utility maximization that are subject to a budget constraint (Arora, Allenby, and Ginter 1998; Chiang 1991; Chintagunta 1993). Assuming a translog bivariate utility function between the product category of interest and the composite good, Chiang (1991) derives a brand-choice model in which the household's no-purchase and brand-choice decisions are shown to be related, in that they share common covariates and parameters. This yields the generalized extreme value (GEV) model, which is identical to the nested logit model discussed previously, except that the coefficient of the inclusive value in the category-purchase model (γ_{2h}) is itself related to the parameters of the brand-choice model (α_{hj}, δ_{2h}, δ_{3h}, and δ_{4h}) (for a detailed exposition of the GEV model, see Maddala 1983).

Although the attractiveness of the GEV model is its utility-theoretic foundations, it suffers from the same limitations as the nested logit model in terms of the absence of correlation between purchase decisions through unobservables and the use of a single-summary measure as the only relevant covariate in the category-incidence model. Furthermore, the GEV model has an additional parametric restriction because it assumes that the coefficient of the inclusive value variable specifically depends on the parameters of the brand-choice model.

Our model is more flexible and general than previously proposed models of category purchase and brand choice in three important ways. First, it models the correlation between the household's category-purchase and brand-choice decisions that may arise from common unobservable influences. Second, it allows for a flexible correlation structure in a household's response sensitivities to marketing variables across purchase decisions. Third, in the category-incidence model, our model includes a general set of covariates, specifically the marketing variables associated with all brands in the product category rather than a summary measure based on the covariates.

DESCRIPTION OF DATA

We obtained data from Information Resources Inc.'s scanner-panel database on household purchases in a metropolitan market in a large U.S. city. For our analysis, we chose the cola product category because the impact of unobserved variables (e.g., temperature, arrival of guests, planning for a party) on households' category-purchase and brand-choice decisions may be quite significant. The data set covers a two-year period from June 1991 to June 1993 and contains shopping-visit information on 494 panelists across four different stores in an urban market. The data set contains information on marketing variables (price, in-store displays, and newspaper feature advertisements) at the stockkeeping unit (SKU) level for each store per week.

The choice of households that purchased at the two largest stores in the market (which collectively account for 90% of all shopping visits in the database) yields 488 households. From these households, we selected a random sample of 350 households and excluded households that never made a purchase in the cola category, for a total of 312 households.[2] These households made a total of 20,236 shopping visits (at the two largest stores), of which 22.5% of the visits resulted in the purchase of cola.

There are four brands in the cola category: Coca-Cola, Pepsi, RC Cola, and a private label. For shopping visits that involved cola purchase, we computed the marketing variables for the nonpurchased brands as share-weighted average values across all SKUs represented by that brand name.[3] For shopping visits that did not involve purchase of any cola brand, we computed the marketing variables of all brands using this share-weighting procedure. Descriptive statistics pertaining to the marketing variables of brands are provided in Table 1. Among the four brands, Pepsi has the highest market share (53%) and is displayed and featured more than other brands. Coca-Cola is the highest-priced brand in the category; the private label is the lowest priced and has the lowest share (9%).

Table 1
DESCRIPTIVE STATISTICS OVER STUDY PERIOD

Brand	*Share*	*Price*	*Display*	*Feature*
Private label	.09	.56	.08	.06
Pepsi	.53	.77	.18	.27
Coca-Cola	.24	.84	.11	.19
RC Cola	.14	.80	.08	.09

In our estimation, as we discussed in the previous section, we include the price, display, and feature values of all brands and the household's product inventory as covariates in the category-purchase model. Price is a continuous variable that is operationalized in dollars per regular package size (which is some unspecified multiple of ounces, given the heterogeneous package sizes in this category),[4] and display and feature are indicator variables that take values between 0 and 1 depending on the percentage of SKUs of that brand that were on display or feature that week. Inventory is a continuous variable (measured in regular package size) that we compute using the household's product consumption rate, which in turn we compute by dividing the total product quantity that the household purchased over the study period by the number of weeks in the data. For the first week in the data, we assume that each household has enough inventory for that week (i.e., we assume that the inventory variable for a household at t = 1 is the household's weekly product consumption rate). Our operationalization of inventory is consistent with previous research on brand-choice models (see, e.g., Chintagunta 1993). We ignore the household's purchase quantity decision and treat a household's multiunit purchase as independent single-unit purchases (e.g., Gupta et al. 1996).

EMPIRICAL RESULTS

We estimate the proposed model of brand choice with a no-purchase option and the nested logit model. To be consistent with the specification of unobserved heterogeneity across households in the proposed model, we assume that the parameters of the nested logit model are also drawn from a multivariate normal distribution.

To verify the predictive power of the proposed model with respect to the nested logit model, we reserve a holdout sample of 112 households and compute predictive marginal log-likelihoods for the purchases in the holdout data. For the proposed and nested logic models, these measures are –5399 and –5545, respectively, and the corresponding in-sample log-likelihoods are –14,473 and –14,802, respectively. We also compute the mean absolute deviation (MAD) between the predicted choice probabilities and the observed choice outcomes at both the category level and the brand level. For the proposed and nested logit models, the MAD measures for category purchase in the holdout sample are .312 and .331, respectively, and the corresponding in-sample measures are .338 and .354, respectively. According to the proposed model, the MAD measures for brand choices in the holdout sample are .150, .375, .297, and .180

[2]Our results largely remained the same when we reincluded these households in the empirical analysis.

[3]We computed share on the basis of number of units sold instead of revenue.

[4]It is useful to note that retail price promotions account for most of the observed variation in the price measure over time.

for the private label, Pepsi, Coca-Cola, and RC Cola, respectively (the in-sample counterparts are .176, .363, .262, and .175, respectively). According to the nested logit model, the corresponding holdout measures are .165, .393, .314, and .208, respectively (the in-sample counterparts are .194, .389, .278, and .202, respectively). Both the log-likelihood measure and the MAD measure show the realized improvements over the nested logit model.[5]

We now report the estimates of the error covariance matrix Ω. The results (see Table 2) show that the estimated correlation between category purchase and brand choice is significantly different from zero for all three national brands (i.e., Pepsi, Coca-Cola, and RC Cola) and equal .22, .42, and .29, respectively. As we mentioned previously, these correlations, which are caused by unobservables affecting both decisions, are large and statistically significant. We do not model the precise sources of the correlations, but we can guess as to the mechanisms by which they arise. If households systematically purchase Coca-Cola whenever they expect guests during a week, this would manifest as correlation between the household's category-purchase decision and its brand-choice decision for Coca-Cola (because the arrival of guests is unobserved and thus relegated to the error terms in the model). National advertising efforts of Pepsi, for example, that are typically unobserved in scanner-panel data may be responsible for inducing correlation in the category incidence and Pepsi brand-choice decisions. Further understanding of the drivers of such correlation is of practical interest to retailers for planning promotional policies for cola with respect to policies for the cola category as a whole. We find that the unobserved correlations are negative (–.35, –.10, and –.20 for Pepsi, Coca-Cola, and RC Cola, respectively) if we ignore correlations in the unobserved heterogeneity distribution between the two decisions. We also find that ignoring the unobserved correlations altogether results in overstating of the estimated variances of brands' utilities (which become 1.37, 1.46, and 1.32 for Pepsi, Coca-Cola, and RC Cola, respectively).

The estimated parameters (specifically the posterior means and standard deviations) of the category-purchase model are presented in Table 3. All the estimated parameters, except the one associated with the price of the private label, have the expected signs. The price of Pepsi has a coefficient of –2.48, compared with coefficients of –.74, –.72, and .18 for Coca-Cola, RC Cola, and the private label,

[5]Estimation of a model that ignores the unobserved correlations in the error terms shows that most of the fit improvement in our proposed model derives from the modeling of flexible coefficients across brands in the no-purchase model, not from the modeling of unobserved correlations between the two decisions.

Table 2

ESTIMATED COVARIANCE MATRIX OF ERROR TERMS: PROPOSED MODEL

Category	*Pepsi*	*Coca-Cola*	*RC Cola*
1	.22 (.07)	.42 (.07)	.29 (.06)
.22 (.07)	1 + .21(.05)	1	1
.42 (.07)	1	1 + .29 (.03)	1
.29 (.06)	1	1	1 + .14 (.06)

Table 3

ESTIMATED PARAMETERS OF THE CATEGORY-PURCHASE MODEL: PROPOSED MODEL

Parameter	*Mean*	*Standard Deviation*
Intercept	1.40	.19
Price		
Private label	.18	.16
Pepsi	–2.48	.22
Coca-Cola	–.74	.14
RC Cola	–.72	.13
Display		
Private label	.15	.09
Pepsi	.44	.06
Coca-Cola	.42	.07
RC Cola	.33	.08
Feature		
Private label	.21	.09
Pepsi	.23	.06
Coca-Cola	.16	.07
RC Cola	.06	.08
Inventory	–.04	.01

respectively. This implies that a retail price cut on Pepsi has a much greater effect than an equal-sized price cut on any other brand, in terms of increasing primary demand for the cola category. Similarly, Pepsi has display and feature coefficients that are larger in magnitude than those associated with other brands.[6] Overall, this suggests that Pepsi is a category cue in that it influences market demand for the cola category. This departs from the restriction inherent in the nested logit model that the prices of all brands have an equal effect (beyond asymmetries that arise from the estimation of different intercepts for brands in the brand-choice model) on the household's category-purchase decision, and it demonstrates the flexibility of the proposed model in explaining households' category-purchase decisions. A managerial implication of this finding is that, all else being equal, the retailer has a greater incentive to promote Pepsi than other brands, because such a promotion increases the market's cola consumption as a whole (i.e., primary demand). Whether such increased consumption is at the expense of another possibly higher-margin product category can be investigated only with a cross-category framework.

Table 4 contains the estimated parameters of the brand-choice portion of our model. The intercepts associated with all brands are positive, which makes sense because the intercepts are estimated relative to the private label. Pepsi is

[6]However, the differences in display and feature coefficients are significant for only the private label and RC Cola, respectively.

Table 4

ESTIMATED PARAMETERS OF THE BRAND-CHOICE MODEL: PROPOSED MODEL

Parameter	*Mean*	*Standard Deviation*
Intercept		
Pepsi	1.50	.09
Coca-Cola	.94	.09
RC Cola	.73	.10
Price	–2.85	.18
Display	.58	.07
Feature	.19	.05

associated with the largest intercept, which suggests that Pepsi has the largest baseline preference in the cola category. The parameter estimates associated with the marketing-mix variables have the expected signs. For example, the estimate of the price parameter is negative, and that of the display or feature parameter is positive. We find that ignoring correlations in the unobserved heterogeneity distribution between the parameters of the no-purchase model and the brand-choice model results in overstating of the estimated brand intercepts (2.35, 1.76, and 1.41 for Pepsi, Coca-Cola, and RC Cola, respectively). However, we also find that ignoring unobserved correlations between households' no-purchase and brand-choice outcomes results in overstating of not only the estimated intercepts for the national brands (1.89, 1.43, and .94 for Pepsi, Coca-Cola, and RC Cola, respectively) but also the estimated marketing-mix coefficients (–3.53, .76, and .27 for price, display, and feature, respectively).

In Table 5, we present a small subset of the significant estimates from the covariance matrix associated with the unobserved heterogeneity distribution (the full covariance matrix involves 210 parameters, of which we estimated that 44 are significant). There is substantial variation across households in the category intercept (posterior mean = 2.67), which represents category usage, and the coefficient associated with Pepsi's price (posterior mean = 6.47). The covariance between Pepsi's price parameter and the category intercept is negative (posterior mean = –3.46), as is the covariance between the price parameter in the brand-choice model and the category intercept (posterior mean = –1.74). This implies that frequent cola buyers are more sensitive both to Pepsi's price in their category-purchase decision and to price in general in their brand-choice decision. To the extent that frequent cola buyers are more likely to be aware of prices in the product category, their heightened responsiveness to prices makes intuitive sense.

The covariance between RC Cola's price parameter and the private label's price parameter is negative (posterior mean = –.60), which implies that households that are more sensitive to RC Cola's price are less sensitive to the private label's price in their category-purchase decision. We were not surprised that households that purchase a national brand (e.g., RC Cola) as a category cue in determining whether to purchase the product category on a given week do not respond to a private label's price for the same purpose. We also find that households with high intrinsic preference for Pepsi also have high intrinsic preference for Coca-Cola (posterior mean for the covariance = .77).

The covariance between Pepsi's price and feature parameter in the category-purchase model and the price and feature parameter in the brand-choice model is positive (price: posterior mean = 3.01; feature: posterior mean = .07), which implies that households that are more sensitive to Pepsi's price and feature in their category-purchase decisions are also more sensitive in general to price and feature in their brand-choice decisions. Existing models of category purchase and brand choice have not estimated such correlations in a household's marketing-mix parameters across purchase decisions. It has been assumed that the correlations are either absent or arbitrarily imposed from the theoretical primitives of the model.

The estimated covariance between any brand intercept in the brand-choice model and the intercept of the category-purchase model is positive (e.g., Pepsi's posterior mean = .74), which implies that buyers of national brands purchase cola more frequently than do buyers of private labels. This implies that national brands enjoy double-jeopardy effects in the cola category in that they not only have higher market share than the private label but also attract more frequent cola drinkers. Although we were not surprised by this, we believe that this finding shows that national brands have much to lose from the encroachment efforts of store brands in this product category.

Table 5

ESTIMATED COVARIANCES OF THE UNOBSERVED HETEROGENEITY DISTRIBUTION

Covariance	*Mean*	*Standard Deviation*
Constant–constant	2.67	.79
Price Pepsi–constant	–3.46	.84
Price Pepsi–price Pepsi	6.47	1.20
Price RC Cola–price private label	–.60	.23
Display Coca-Cola–price Coca-Cola	–.31	.12
Feature private label–price private label	.21	.12
Pepsi intercept–constant	.91	.24
Coca-Cola intercept–constant	.74	.22
Coca-Cola intercept–Pepsi intercept	.77	.13
RC Cola intercept–constant	.68	.22
Price–constant	–1.74	.54
Price–price Pepsi	3.01	.63
Feature–feature Pepsi	.07	.03

We find that households that respond more to Coca-Cola's price also tend to respond more to Coca-Cola's display in their category-purchase decisions (posterior mean of the correlation between the coefficients for Coca-Cola's price and display = –.31). In contrast, we find that households that respond more to the private label's price tend not to respond to the private label's feature advertising in their category-purchase decisions (posterior mean of the correlation between the coefficients for the private label's price and display = .21). Overall, therefore, our study is able to recover a highly flexible correlation structure in a household's responsiveness to different marketing variables not only within a given purchase decision (i.e., category purchase or brand choice) but also across purchase decisions.

In Table 6, we report the marketing-mix elasticities from our fitted model. We computed these as follows: For each household, in the period after the estimation period, we assume that there is a specific change in the marketing environment (e.g., a 20% price cut for the price–elasticity computation), and we assess the impact of the change on each household's response outcome by sampling the predictive distribution of responses (see, e.g., Montgomery and Bradlow 1999). Each row in Table 6 corresponds to a specific brand's marketing variable that we varied to assess the response of primary demand (i.e., category-purchase probability) and secondary demand (i.e., brand-choice probabilities), as represented in the columns. Consistent with the parameter estimates in Table 3, the marketing-mix elasticities of category demand are the highest for Pepsi's marketing-mix variables (e.g., the elasticity of category demand in response to Pepsi's price is –2.31, compared with –.83 in response to Coca-Cola's price). Among the brand-choice elasticities, RC Cola has the highest own-price elasticity, and the private label has the lowest.

Table 6
MARKETING-MIX ELASTICITIES: PROPOSED MODEL

Marketing Variable	*Category*	*Private Label*	*Pepsi*	*Coca-Cola*	*RC Cola*
Price					
Private label	.08	–1.05	.22	.06	–.09
Pepsi	–2.31	1.52	–1.50	2.07	2.15
Coca-Cola	–.83	.71	.80	–2.88	1.03
RC Cola	–.72	–.17	.55	.68	–4.67
Display					
Private label	.01	.08	–.02	–.01	.01
Pepsi	.10	–.08	.10	–.14	–.17
Coca-Cola	.05	–.02	–.02	.08	–.03
RC Cola	.03	–.01	–.01	–.01	.07
Feature					
Private label	.01	.03	.00	–.01	–.01
Pepsi	.09	–.04	.03	–.05	–.03
Coca-Cola	.03	–.01	–.02	.08	–.02
RC Cola	.01	–.02	–.00	–.01	.06

Table 7
MARKETING-MIX ELASTICITIES: NESTED LOGIT MODEL

Marketing Variable	*Category*	*Private Label*	*Pepsi*	*Coca-Cola*	*RC Cola*
Price					
Private label	–.26	–2.22	.32	.36	.54
Pepsi	–1.64	1.79	–1.55	2.48	1.84
Coca-Cola	–.62	.68	.84	–3.70	.72
RC Cola	–.36	.63	.38	.44	–3.40
Display					
Private label	.01	.05	–.01	–.01	–.01
Pepsi	.13	–.13	.12	–.20	–.14
Coca-Cola	.03	–.03	–.04	.18	–.03
RC Cola	.01	–.01	–.01	–.01	.07
Feature					
Private label	.00	.03	–.01	–.01	–.01
Pepsi	.06	–.07	.06	–.09	–.07
Coca-Cola	.01	–.02	–.02	.08	–.02
RC Cola	.01	–.01	–.01	–.01	.04

The marketing-mix elasticities for the nested logit model are shown in Table 7. The elasticities of primary demand (i.e., category purchases) with respect to brands' prices are systematically understated in the nested logic compared with the proposed model. The effects of the private label's price on other brands' demand are overstated in the nested logit compared with the proposed model. These findings imply that existing models understate category-expansion effects of price cuts and overstate the effects of the private label's price cuts on national brands' sales.

Percentage decompositions of all brands' (own) marketing-mix elasticities between category purchase and brand choice are presented in Table 8. For the proposed model, averaging across the four brands, we find that category purchase accounts for 21.8% of a brand's total price elasticity, 32% of a brand's total display elasticity, and 35.5% of a brand's total feature elasticity. According to the nested logit model, the corresponding percentages are 21.5, 21.5, and 21.2, respectively. This suggests that currently used models, such as the nested logit model, may severely understate the category-expansion effects of sales promotions. This finding has important implications for retailers that currently use feature activities to induce consumers to switch to higher-margin brands, in that it alerts them to the

Table 8
DECOMPOSITION OF MARKETING-MIX ELASTICITIES (PERCENTAGE OF TOTAL ELASTICITY ACCOUNTED FOR BY CATEGORY PURCHASE)

Marketing Variable	*Proposed Model*	*Nested Logit Model*	*Nested Version of Proposed Model*
Price	21.8	21.5	22.2
Display	32.0	21.5	33.8
Feature	35.5	21.2	38.0

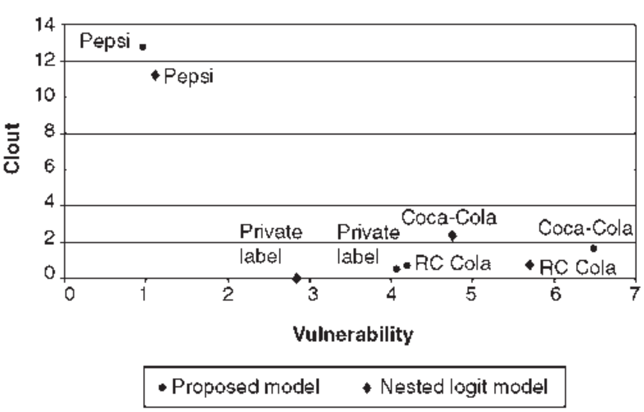
category-expansion benefits of sales promotions. In the rightmost column of Table 8, we report the elasticity decomposition for the nested version of our model that ignores unobserved correlations between the no-purchase and brand-choice decisions. Notably, we find that the elasticity decomposition in the nested version of our model is quite similar to that obtained in the proposed model, despite our finding that all the marketing-mix coefficients are overstated in the nested version.

Because of Bell, Chiang, and Padmanabhan's (1999) findings that category purchase accounts for 7% of the combined category and brand components of a brand's price elasticity in the soft drinks category, we were surprised to find that category purchase accounts for 21.8%. A reason for the difference might be that Bell, Chiang, and Padmanabhan use data for all soft drinks (not just colas), whereas we rely on data for colas only. Therefore, category purchase being responsible for a negligible percentage of the total price elasticity for noncola brands or households' brand-switching between colas and noncolas being substantial would explain the differences in decomposition-related findings between our study and that of Bell, Chiang, and Padmanabhan. It is also useful to note that decomposition of a brand's elasticity measure, though mathematically correct, is substantively uninterpretable. As Van Heerde, Gupta, and Wittink (2003) show, decomposition of a brand's unit sales is managerially more meaningful. Because we ignore the household's purchase-quantity decision, we are unable to perform a decomposition of a brand's unit sales with our model.

On the basis of cross-price elasticities we obtained from the proposed and nested logit models (see Tables 6–7), we compute competitive clout and vulnerability measures for all brands, as do Kamakura and Russell (1989). We compute the competitive clout measure for a brand as the sum of squares of all cross-elasticities in the row of the price-elasticity matrix that corresponds to the price change on that brand. This measures the ability of a brand to affect demand for other brands. We compute the competitive vulnerability measure for a brand as the sum of squares of all cross-elasticities in the column of the price-elasticity matrix that corresponds to the demand response of that brand. This measures the sensitivity of a brand's demand to price changes of other brands. The competitive clout and vulnerability measures in the two model specifications are displayed in Figure 1. In all model specifications, we found that Pepsi has the highest clout and the lowest vulnerability in the category. This is consistent with the accepted idea in grocery retailing that Pepsi is the strongest cola brand in supermarkets. Although the nested logit model estimates the clout and vulnerability measures of the largest and smallest brands fairly accurately (i.e., approximately the same as the proposed model), it overestimates the vulnerability of Coca-Cola and understates the vulnerability of RC Cola. The average distances of estimated positions of brands in the proposed model are shorter than the average distances in the nested logit model.

Figure 1

CLOUT AND VULNERABILITY



Generalizability of Results

To investigate the importance of our proposed modeling framework for other product categories, we reestimated our proposed model using scanner-panel data on coffee and sugar. Unlike soft drinks, for which there are multiple package types (e.g., cans, bottles), these categories are homogeneous in terms of package types. Furthermore, multiunit purchasing is rarely observed. We comment here on the nature of the substantive findings we obtained from these additional data sets.[7]

For coffee, for the proposed and nested logit models, the predictive log-likelihood measures are –14,981 and –15,465, respectively, and the corresponding in-sample log-likelihoods are –36,457 and –37,519, respectively. For sugar, the predictive log-likelihoods are –14,012 and –14,098, and the in-sample measures are –35,717 and –35,844. These findings further support the empirical superiority of the proposed model over the nested logit model.

The estimated correlation between category purchase and brand choice is positive and significant for two of the four coffee brands (i.e., Maxwell House and other) and is equal to .31 and .66, respectively. The correlation is negative and significant for one of the four sugar brands (i.e., American Crystal) and is equal to –.33. The negative value for the estimated correlation is notable in that it implies that when households buy sugar for unobserved reasons, they systematically tend to avoid American Crystal. Unlike soft drinks, in neither coffee nor sugar does the largest share brand (Hills Brothers and Domino) have significant estimates of these correlations. As in soft drinks, there is considerable asymmetry in terms of how various brands' marketing variables influence the household's utility for the no-purchase option. In these categories, we again find that ignoring correlations in the unobserved heterogeneity distribution between the parameters of the no-purchase model and the brand-choice model results in overstating of the estimated brand intercepts in the brand-choice model. We also found that the nested logit understates the category-purchase effects of promotional activities compared with the proposed model.

CONCLUSIONS

In this article, we propose a flexible way to model the category-purchase and brand-choice decisions of households jointly. The proposed model is more general than existing models of the two decisions in three important ways: (1) it models correlations in the unobservable drivers of the category-purchase and brand-choice decisions; (2) it specifies the effects of covariates in the category-purchase model more flexibly by freely estimating the coefficients associated with all brands' marketing variables in the category-purchase model; and (3) it models correlations in a

[7]Detailed results are available from the authors on request.

household's sensitivity to marketing variables between the category-purchase and brand-choice decisions, which enables us to investigate, for example, whether a household that is more responsive to price in its category-purchase decision is also more responsive to price in its brand-choice decision. We estimate the proposed model using Bayesian MCMC methods. We investigate the empirical consequences of ignoring the three modeling innovations inherent in our proposed model by comparing the empirical results we obtained from the proposed model with those we obtained from the widely used nested logit model.

Our key empirical findings are as follows: First, we find that our proposed model fares better, in a predictive sense, than the nested logit model. Second, we find that unobserved drivers of category purchase and brand choice are positively correlated for the three national brands in the category, which implies that when households buy cola for reasons unknown to the researcher, they predominantly tend to buy national brands. Third, we find that Pepsi serves as a category cue in terms of influencing market demand for cola, in that the coefficients associated with its marketing mix are much higher than those associated with the other brands in the category-purchase model. Fourth, we find that frequent cola buyers are more sensitive both to Pepsi's price in their category-purchase decision and to price in general in their brand-choice decision. Fifth, we find that households that are more responsive to Pepsi's price and feature in their category-purchase decision are more sensitive to price and feature in general in their brand-choice decision. Sixth, we find that the nested logit model understates category-expansion effects of price cuts but overstates the effects of a private-label price cut on the sales of the national brands. Seventh, we find that from an elasticity-decomposition standpoint, the nested logit model severely understates the category-purchase component of all marketing-mix elasticities. Last, we estimate the clout and vulnerability measures of the various brands and show that those from the nested logit model appear to be distorted. We discuss the managerial implications of all our empirical findings.

Given the generality and flexibility of our proposed model and the strength of the empirical findings we obtained by calibrating it on scanner-panel data, we believe that the proposed model will prove useful in marketing. Even though models such as the nested logit and the GEV are simpler to estimate, we believe that given the currently available computational resources and degrees of freedom that are inherent in scanner-panel data, there is no need to limit attention to restrictive models unless available data sets warrant such restrictions.[8] It is useful to note that even though our application includes the price, display, and feature covariates at both stages (i.e., category purchase and brand choice) of the econometric model, our framework is applicable to cases in which the covariates are not identical between the two stages.[9]

Several notable problems can be taken up in further research. First, it would be of interest to understand explicitly the drivers of the estimated correlations between category-purchase and brand-choice decisions through supplementary data sources, such as national advertising expenditures of brands and household surveys on brand attitudes and preferences. A way to achieve this would be to allow Σ in our model to be a function of observable covariates.[10] Second, given the advent of basket-level data, it also would be of interest to model jointly and estimate cross-category dependence in category-purchase and brand-choice decisions. This is a problem of significant computational complexity, yet it is necessary to exploit fully the scale and scope of available basket data for marketing purposes, especially given the increasing interest in the modeling of cross-category-purchase behavior of households. We believe that our modeling and analytical framework will facilitate the construction of a comprehensive cross-category model of purchase incidence and brand choice. For example, allowing household indirect utility for the no-purchase option and for specific brands to be related across categories is a way to flexibly capture cross-category dependence in household purchase decisions due to, for example, consumption complementarity, budget constraint, and umbrella-branding effects. Third, as Van Heerde, Leeflang, and Wittink's (2003) recent study shows, using store-level data, price promotions have different effects on brand sales depending on whether they are accompanied by display and/or feature support; thus, it would be useful to understand whether such interaction effects characterize household-level brand-choice data and, if so, whether our estimated unobserved correlations are partly due to such interactions. Fourth, we use a household-specific share-weighted average across relevant SKUs to compute the price variable at the brand level. Although this conforms to popular practice in the brand-choice literature, it would be useful to investigate the impact of other aggregation schemes (e.g., the minimum price among relevant SKUs) on the inferences obtained. Fifth, and similar to the previous point, it would be of interest to investigate whether different package types and/or sizes must be analyzed separately rather than aggregated to the brand level. Finally, it would be of interest to combine a model of purchase quantity with our proposed model of category purchase and brand choice and possibly to model inventory-based dynamics in households' purchases (see, e.g., Erdem, Imai, and Keane 2003).

APPENDIX: ESTIMATION PROCEDURE

The objective of the analysis is to estimate the parameters $\psi = (\beta, D, \omega)$, where ω consists of the $2 \times (J - 1)$ free parameters in the covariance matrix Ω. The observed purchase-incidence data on the hth household are denoted as $y_h = (y_{h1}, ..., y_{hn_h})$, and the data on the brand choices are denoted as $y_h^* = (y_{h1}^*, ..., y_{hn_h}^*)$, where y_{ht}^* is 1, 2, ..., J if $y_{ht} = 1$ or is not observed otherwise. In addition, $y = (y_1, ..., y_n)$ and $y^* = (y_1^*, ..., y_n^*)$ denote the complete set of responses on all n households in the panel. For the covariates, we define X_{ht} as

[8]However, because the proposed model is more heavily parameterized than existing models, researchers should not judge the empirical superiority of the proposed model on the basis of traditional in-sample fit measures without adjusting for the additional number of parameters.

[9]We thank an anonymous reviewer for alerting us to this issue.

[10]We thank an anonymous reviewer for this suggestion.

$$\text{(A1)}\quad \begin{pmatrix} 1 & P_{ht1} & \cdots & P_{htJ} & D_{ht1} & \cdots & D_{htJ} & F_{ht1} & \cdots & F_{htJ} & I_{ht} & 0 & \cdots & 0 & 0 & 0 & 0 \\ 0 & 0 & \cdots & 0 & 0 & \cdots & 0 & 0 & \cdots & 0 & 0 & 1 & \cdots & 0 & P^{\dagger}_{ht2} & D^{\dagger}_{ht2} & F^{\dagger}_{ht2} \\ \vdots & \vdots & & & & & & & & & & & & & \vdots & \vdots & \vdots \\ 0 & 0 & \cdots & 0 & 0 & \cdots & 0 & 0 & \cdots & 0 & 0 & 0 & \cdots & 1 & P^{\dagger}_{htJ} & D^{\dagger}_{htJ} & F^{\dagger}_{htJ} \end{pmatrix},$$

which is a $J \times k$ matrix, where k is the total number of covariates in the model, including intercepts ($k = 1 + 3 \times J + 1 + J - 1 + 3 = 4[J + 1]$). In addition, X_h is the matrix $(X'_{h1}, X'_{h2}, ..., X'_{hn_h})'$, $\mathbf{z}'_{ht} = (\mathbf{z}_{ht}, u^{\dagger}_{ht2}, ..., u^{\dagger}_{htJ})'$ denotes the latent random utility of category purchase and brand choice of household h at time t, and $\mathbf{z}_h = (\mathbf{z}'_{h1}, ..., \mathbf{z}'_{hn_h})'$ denotes the latent utilities across n_h shopping occasions. Under our assumptions, it follows that

$$\text{(A2)}\qquad \mathbf{z}_{ht}|\beta_h, \psi \sim N(X_{ht}\beta_h, \Omega^{\dagger}).$$

Conditioned on the random effects β_h, it follows that the probability of the outcome (y_{ht}, y^*_{ht}) is

$$\text{(A3)}\qquad \Pr(y_{ht}, y^*_{ht}|\beta_h, \psi) = \int_{A_{ht}} \int_{B_{ht2}} \cdots \int_{B_{htJ}} \phi(\mathbf{z}_{ht}|X_{ht}\beta_h, \Omega^{\dagger})dz_{ht},$$

and when $y_{ht} = 1$ and $y^*_{ht} = j\ (j > 1)$,

$$\text{(A4)}\qquad A_{ht} = (0, \infty),$$

$$\text{(A5)}\qquad B_{htj} = \left(\max\left\{0, \max_{k \neq j} u^{\dagger}_{htk}\right\}, \infty\right), \text{ and}$$

$$\text{(A6)}\qquad B_{htk} = (-\infty, u^{\dagger}_{htj}),\ k \neq j.$$

When $y_{ht} = 1$ and $y^*_{ht} = 1$,

$$\text{(A7)}\qquad A_{ht} = (0, \infty), \text{ and}$$

$$\text{(A8)}\qquad B_{htk} = (-\infty, 0),\ k > 1.$$

When the no-purchase outcome occurs, $A_{ht} = (-\infty, 0)$, and B_{htk} is the entire real line. We can formally define the probability of the sequence of outcomes on the hth household as follows:

$$\text{(A9)}\qquad \Pr(y_h, y^*_h|\psi) = \int \prod_{t=1}^{n_h} \Pr(y_{ht}, y^*_{ht}|\beta_h, \psi)\phi(\beta_h|\beta, D)d\beta_h,$$

and the joint probability across households is given by

$$\text{(A10)}\qquad \Pr(y, y^*|\psi) = \prod_{h=1}^{H} \Pr(y_h, y^*_h|\psi).$$

We readily observe that the joint probability function (the likelihood function of the parameters, given the data) is intractable, which makes it difficult for us to obtain the maximum-likelihood estimate, even using simulation methods. Therefore, we adopt a simulation-based Bayesian approach to estimate the model parameters; the approach bypasses the computation of the intractable likelihood function. In this approach, we enlarge the parameter space to include the latent $\{z_h\}$ and $\{\beta_h\}$ in the sampling (Tanner and Wong 1987), and we apply the methods of Albert and Chib (1993) and Chib and Greenberg (1998) to sample the various unknowns in blocks, conditioned on the most current values of the remaining blocks.

Prior Distributions

Because our approach to estimation is Bayesian, we complete the model formulation by specifying prior distributions for the unknown model parameters. In particular, we assume that

$$\text{(A11)}\qquad \beta \sim N_k(\beta_0, B_0),$$

$$\text{(A12)}\qquad \omega \sim N_{2 \times (J-1)}(g_0, G_0), \text{ and}$$

$$\text{(A13)}\qquad D^{-1} \sim \text{Wishart}_k(\rho_0, R_0),$$

where we define the parameters as follows: β_0 is a k-dimensional vector of zeros, B_0 is a $k \times k$ diagonal matrix that has its diagonal elements set equal to 10, g_0 is a $2 \times (J - 1)$–dimensional vector with elements that correspond to variances set equal to 1 and elements that correspond to covariances set equal to 0, G_0 is a $(2 \times [J - 1]) \times (2 \times [J - 1])$ identity matrix, $\rho_0 = k + 4$, and R_0 is a $(k \times k)$–dimensional identity matrix. We chose this prior specification to allow the data to dominate the results.

MCMC Algorithm

The main idea of the estimation approach is to focus on the posterior distribution of the parameters, the latent data, and the random effects and then to summarize this posterior distribution with MCMC methods (Chib 2001; Chib and Greenberg 1996; Tierney 1994). Using MCMC methods, we design an ergodic Markov chain with the property that the limiting invariant distribution of the chain is the posterior density of interest. Then, we can take draws furnished by sampling the Markov chain, after an initial transient or burn-in stage, as approximate correlated draws from the posterior distribution. This output forms the basis for a summary of the posterior distribution and for computation of Bayesian point and interval estimates. Ergodic laws of large numbers for Markov chains on continuous state spaces are used to justify that the estimates are simulation consistent and converge to the posterior expectations as the simulation sample size becomes large.

A standard method for constructing a Markov chain with the correct limiting distribution is through a recursive simulation of the so-called full conditional densities (i.e., the density of a set or block of parameters), given the data and the remaining blocks of parameters. Each full conditional density in the simulation is then sampled either directly (if the full conditional density belongs to a known family of distributions) or through a technique such as the Metropolis–Hastings method (Chib and Greenberg 1995).

In the present case, the posterior distribution of interest is defined as follows:

(A14)
$$\pi(\{\beta_h\}, \{z_{ht}\}, \beta, D, \omega | y, y^*) \propto \pi(\beta)\pi(D)\pi(\omega)\prod_{h=1}^{H}\prod_{t=1}^{n_h}\phi(z_{ht}|X_{ht}\beta_h, \Omega^\dagger),$$

where π denotes the prior and posterior densities. To sample the posterior distribution, we rely primarily on the techniques introduced by Albert and Chib (1993) and Chib and Greenberg (1998). We constructed our Markov chain using the following full conditional distributions:

- $\beta_h|z_h, \psi, h = 1, 2, ..., H;$
- $z_{ht}|y, y^*, \{\beta_h\}, \psi, \{z_{(-ht)}\}, t = 1, ..., n_h; h = 1, 2, ..., H;$
- $u^\dagger_{htk}|y, y^*, \{\beta_h\}, \psi, \{z_{ht}\}, \{u^\dagger_{(-htk)}\}, k = 2, ..., J; t = 1, ..., n_h; h = 1, 2, ..., H;$
- $\beta|\{\beta_h\}, D;$
- $D|\{\beta_h\}, \beta;$ and
- $\omega|\{z_h\}, \{\beta_h\};$

where, for example, $\{z_{(-ht)}\}$ denotes the entire set of z's, excluding z_{ht}. After specifying some realistic but arbitrary starting values, we simulated the conditional distributions many times, and we used the most current values of the conditioning variables in each simulation. We iterate this sampling procedure to produce 11,000 draws, of which we discarded the first 1000 to remove the effect of the initial conditions. According to the theory of MCMC sampling, the draws constitute a sample (albeit correlated) from the joint posterior distribution $\pi(\{\beta_h\}, \{z_{ht}\}, \beta, D, \omega|y, y^*)$.

Next, we provide more details on the particular form of each of the full conditional distributions. The first five distributions are tractable and are sampled directly; the last distribution is sampled by the Metropolis–Hastings algorithm (for details on how this is done in a similar application, see Chib and Greenberg 1998; Chib, Seetharaman, and Strijnev 2002).

It is not difficult to check (by standard Bayesian calculations) that the distribution $\beta_h|y, \beta, z_h, \omega, D$ is $N_k(\hat{\beta}_h, B_h)$, where $B_h = [D^{-1} + X'_h(I_{n_h} \otimes \Omega^{-1})X_h]^{-1}$, and $\hat{\beta}_h = B_h[D^{-1}\beta + X'_h(I_{n_h} \otimes \Omega^{-1})z_h]$. Note that I_{n_h} is an $n_h \times n_h$ identity matrix. Next, it follows from the work of Albert and Chib (1993) that the distribution of z_{ht} is univariate truncated normal with support $(0, \infty)$ if $y_{ht} = 1$ and support $(-\infty, 0)$ if $y_{ht} = 0$. We obtained the parameters of this univariate normal distribution by beginning with the representation $z_{ht}|\beta_h, \psi \sim N(X_{ht}\beta_h, \Omega^\dagger)$ and computing the conditional distribution of the first element (i.e., z_{ht}) using the usual multivariate normal theory. We computed the distribution of $u^\dagger_{htk}$, conditioned on the remaining elements, in the same way, except that the truncation is a bit more involved. When $y_{ht} = 1$ and $y^*_{ht} = j$, with $j > 1$ (i.e., the base brand is not chosen), $u^\dagger_{htj}$ is sampled from a univariate truncated normal distribution with support $(\max\{0, \max_{k \neq j} u^\dagger_{htk}\}, \infty)$, and the remaining brands are sampled from a univariate truncated normal with support $(-\infty, u^\dagger_{htj})$, where $u^\dagger_{htj}$ is the sampled value for the chosen brand. When $y_{ht} = 1$ and $y^*_{ht} = 1$, then each $u^\dagger_{htk}$ is sampled from conditional normal distributions with support $(-\infty, 0)$. Finally, when $y_{ht} = 0$, each $u^\dagger_{htk}$ is sampled from conditional normal distributions without any restriction on the support.

Next, we can easily confirm that the distribution $\beta|y, \beta_h, D$ is $N_k(\hat{\beta}, \hat{B})$, where $\hat{B} = (B_0^{-1} + \sum_h \sum_t D^{-1})^{-1}$, and $\hat{\beta} = \hat{B}(B_0^{-1}\beta_0 + \sum_h \sum_t D^{-1}\beta_h)$. By similar calculations, the distribution $D^{-1}|y, \beta, \beta_h$ is Wishart $W_k(\rho_0 + H, R)$, where $R = [R_0^{-1} + \sum_h(\beta_h - \beta)(\beta_h - \beta)]^{-1}$.

The last distribution, that of ω, is proportional to

(A15)
$$N_{2\times(J-1)}(g_0, G_0)\prod_h\prod_t N_J(z_{ht}|X_{ht}\beta_h, \Omega^\dagger),$$

restricted to the region that generates a positive definite covariance matrix $\Omega^\dagger$. This distribution is not of known form and is sampled by the Metropolis–Hastings algorithm. Specifically, the Metropolis–Hastings step is implemented by the method of tailoring, which Chib and Greenberg (1995) propose, wherein the proposal distribution is taken to be multivariate t with the parameters found from the mode and Hessian of the log of the density shown previously. Tailoring ensures that the acceptance rates are high and the inefficiency factors are low. Inefficiency factors are a measure of the serial correlation of the chain (for a discussion, see Chib 2001).

REFERENCES

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

Albert, James and Siddhartha Chib (1993), "Bayesian Analysis of Binary and Polychotomous Response Data," *Journal of the American Statistical Association*, 88 (June), 669–779.

Arora, Neeraj, Greg M. Allenby, and James L. Ginter (1998), "A Hierarchical Bayes Model of Primary and Secondary Demand," *Marketing Science*, 17 (1), 29–44.

Bass, Frank M. (1974), "The Theory of Stochastic Preference and Brand Switching," *Journal of Marketing Research*, 11 (May), 1–20.

Bell, David R., Jeongwen Chiang, and V. Padmanabhan (1999), "The Decomposition of Promotional Response: An Empirical Generalization," *Marketing Science*, 18 (4), 504–526.

Bucklin, Randolph E. and Sunil Gupta (1992), "Brand Choice, Purchase Incidence, and Segmentation: An Integrated Modeling Approach," *Journal of Marketing Research*, 29 (May), 201–215.

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

Chib, Siddhartha (2001), "Markov Chain Monte Carlo Methods: Computation and Inference," in *Handbook of Econometrics*, Vol. 5, J.J. Heckman and E. Leamer, eds. Amsterdam: North Holland, 3569–3649.

——— and Edward Greenberg (1995), "Understanding the Metropolis-Hastings Algorithm," *American Statistician*, 49 (November), 327–35.

——— and ——— (1996), "Markov Chain Monte Carlo Simulation Methods in Econometrics," *Econometric Theory*, 12 (August), 409–431.

——— and ——— (1998), "Analysis of Multivariate Probit Models," *Biometrika*, 85 (June), 347–61.

———, P.B. Seetharaman, and Andrei Strijnev (2002), "Analysis of Multicategory Purchase Incidence Decisions Using IRI Market Basket Data," *Advances in Econometrics*, 16, 57–92.

Chintagunta, Pradeep K. (1993), "Investigating Purchase Incidence, Brand Choice, and Purchase Quantity Decisions of Households," *Marketing Science*, 12 (1), 184–208.

——— (2002), "Investigating Category Pricing Behavior in a Retail Chain," *Journal of Marketing Research*, 39 (May), 141–54.

———, Dipak C. Jain, and Naufel J. Vilcassim (1991), "Investigating Heterogeneity in Brand Preferences in Logit Models for Panel Data," *Journal of Marketing Research*, 28 (November), 417–28.

Erdem, Tülin, Susumu Imai, and Michael P. Keane (2003), "A Model of Consumer Brand and Quantity Choice Dynamics Under Price Uncertainty," *Quantitative Marketing and Economics*, 1 (1), 5–64.

Gönül, Füsun and Kannan Srinivasan (1993), "Modeling Multiple Sources of Heterogeneity in Multinomial Logit Models: Methodological and Managerial Issues," *Marketing Science*, 12 (3), 213–29.

Guadagni, Peter M. and John D.C. Little (1983), "A Logit Model of Brand Choice Calibrated on Scanner Data," *Marketing Science*, 2 (3), 203–238.

Gupta, Sachin, Pradeep K. Chintagunta, Anil Kaul, and Dick R. Wittink (1996), "Do Household Scanner Data Provide Representative Inferences from Brand Choices: A Comparison with Store Data," *Journal of Marketing Research*, 33 (November), 383–98.

Hanemann, M.W. (1991), "Markov Chain Monte Carlo Simulation Methods in Econometrics," *Econometric Theory*, 12, 409–431.

Kamakura, Wagner A. and Gary J. Russell (1989), "A Probabilistic Choice Model for Market Segmentation and Elasticity Structure," *Journal of Marketing Research*, 26 (November), 379–90.

Little, Roderick J.A. and Donald B. Rubin (1987), *Statistical Analysis with Missing Data*. New York: John Wiley & Sons.

Maddala, G.S. (1983), *Limited Dependent and Quantitative Variables in Econometrics*. Cambridge, UK: Cambridge University Press.

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

Massy, William F., David B. Montgomery, and Donald G. Morrison (1970), *Stochastic Models of Buying Behavior*. Cambridge: Massachusetts Institute of Technology Press.

Montgomery, Alan L. and Eric T. Bradlow (1999), "Why Analyst Overconfidence About the Functional Form of Demand Can Lead to Overpricing," *Marketing Science*, 18 (4), 569–83.

Rossi, Peter E. and Greg M. Allenby (1993), "A Bayesian Approach to Estimating Household Parameters," *Journal of Marketing Research*, 30 (May), 171–82.

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

Tanner, M. and W. Wong (1987), "The Calculation of Posterior Distributions by Data Augmentation," *Journal of the American Statistical Association*, 82 (June), 528–49.

Tierney, Luke (1994), "Markov Chains for Exploring Posterior Distributions (with discussion)," *Annals of Statistics*, 22 (4), 1701–1762.

Van Heerde, Harald J., Sachin Gupta, and Dick R. Wittink (2003), "Is 75% of the Sales Promotion Bump Due to Brand Switching? No, Only 33% Is," *Journal of Marketing Research*, 40 (November), 481–91.

———, Peter S.H. Leeflang, and Dick R. Wittink (2003), "Flexible Unit-Sales Decompositions of Price Promotion Effects Based on Store Data," working paper, Faculty of Economics, Tilburg University.