ANALYSIS OF MULTI-CATEGORY PURCHASE INCIDENCE DECISIONS USING IRI MARKET BASKET DATA

Siddhartha Chib, P. B. Seetharaman and Andrei Strijnev

ABSTRACT

Empirical studies in Marketing have typically characterized a household's purchase incidence *decision, i.e. the household's decision of whether or not to buy a product on a given shopping visit, as being independent of the household's purchase incidence decisions in other product categories. These decisions, however, tend to be related both because product categories serve as complements (e.g. bacon and eggs) or substitutes (e.g. colas and orange juices) in addressing the household's consumption needs, and because product categories vie with each other in attracting the household's limited shopping budget. Existing empirical studies have either ignored such inter-relationships altogether or have accounted for them in a limited way by modeling household purchases in* pairs *of complementary product categories. Given the recent availability of IRI market basket data, which tracks purchases of panelists in several product categories over time, and the new computational Bayesian methods developed in Albert and Chib (1993) and Chib and Greenberg (1998), estimating high-dimensional multi-category models is now possible. This paper exploits these developments to fit an appropriate panel data multivariate probit model to household-level contemporaneous purchases in twelve product categories, with the descriptive goal of isolating*

Econometric Models in Marketing, Volume 16, pages 57–92.

ISBN: 0-7623-0857-5

correlations amongst various product categories within the household's shopping basket. We provide an empirical scheme to endogenously determine the degree of complementarity and substitutability among product categories within a household's shopping basket, providing full details of the methodology. Our main findings are that existing purchase incidence models underestimate the magnitude of cross-category correlations and overestimate the effectiveness of the marketing mix, and that ignoring unobserved heterogeneity across households overestimates cross-category correlations and underestimate the effectiveness of the marketing mix.

1. MOTIVATION

Over the past decade, marketing researchers have devoted a lot of attention to the problem of modeling household purchase incidence at the category level (see, for example, Chiang, 1991; Bucklin & Lattin, 1991; Chintagunta, 1993). One reason for modeling category purchase incidence, in addition to brand-choices within the product category, is that such a model provides improved estimates of brand-choice elasticities with respect to marketing mix variables, properly accounting for not just the direct impact but also the indirect impact on brand-choice via category purchase incidence (Chiang, 1991). A second reason stems from the researcher's desire to understand what factors drive category purchase incidence and what impact, if any, marketing-mix variables at the brand level have on category purchase incidence. A third reason is the purely descriptive goal of isolating correlations amongst various product categories within the household's shopping basket, thereby providing a scheme to determine which categories are complements and which are substitutes.

Previous studies have largely focused on the first issue, i.e. obtaining improved estimates of brand-choice elasticities. The second issue, i.e. estimating the impact of brands' marketing variables on category purchase incidence, and the third issue, i.e. estimating cross-category correlations, have been incompletely addressed at best. While the former is in part due to the difficulty of formulating appropriate models of category purchase incidence, the latter is largely due to the computational problems of fitting realistic household-level category purchase incidence models on scanner panel data. For example, if a household buys thirty different product categories during a visit to the store, a model that estimates cross-category correlations must simultaneously model household decisions in thirty different product categories, an onerous task by any standards. The purpose of this paper, which is part of a two-stage research agenda, is to explicitly address the third issue, i.e. estimate cross-category correlations within the household's shopping basket. We study what information is contained in category purchase incidence data when, not just two or three, but a large number of category purchase incidence decisions (twelve in our case) are modeled simultaneously. The success of our fitting enterprise, based on the work of Albert and Chib (1993) and Chib and Greenberg (1998), and summarized in this paper, makes us hopeful that we will be able to scale-up our model to include all the twenty or so categories in the typical shopping basket. The second-stage of our research, described in a companion paper, addresses all three issues simultaneously, i.e. jointly modeling category purchase incidence and brand choice when the number of categories is large.

2. OBJECTIVES OF THIS STUDY

Households make purchase decisions in several product categories when they visit the supermarket. For example, a household's regularly scheduled trip to the grocery store may involve the purchase of soft drinks, chips, ketchup, cookies, peanut butter, ice cream, laundry detergents, etc. To the extent that product categories serve different consumption needs of the household, household purchase decisions may appear to be independent across product categories within the household's shopping basket. For example, a household's decision to purchase laundry detergents may be independent of the household's decision to purchase bacon or soft drinks since each product serves a fundamentally different consumption need. On the basis of this independence assumption, empirical researchers typically estimate household purchase incidence decisions separately for each product category, i.e. whether or not a household will buy ketchup during a visit to the store is modeled independently of whether or not it will purchase other products in the store (see, for example, Bucklin & Lattin, 1991; Chiang, 1991). This is also referred to as the *weak separability* assumption.

It is unlikely that the weak separability assumption applies to *all* product categories within a household's shopping basket. For example, some products may serve as consumption complements of each other (say, bacon and eggs) while others may serve as consumption substitutes of each other (say, cola and orange juice).[1] Researchers have accounted for this by identifying pairs of products, a priori, that are obvious complements of each other and estimating bivariate models of household purchase incidence decisions across the two product categories (Chintagunta & Haldar, 1998; Manchanda, Ansari & Gupta, 1999). Such a framework is applicable only when one can identify a priori relationships among product categories. In general, however, one must *endogenously* infer the relationships between product categories within the household's shopping basket using purchase data. For example, one must estimate a high-dimensional model of household purchase incidence decisions across all product categories within the household's shopping basket (also referred to as a *basket-level model* henceforth). Such a basket-level model will endogenously estimate correlations across all pairs of product categories rather than across predefined product categories only. Even if the focus is on estimating correlations among and/or marketing mix elasticities within predefined pairs of product categories (as in Chintagunta & Haldar, 1998; Manchanda, Ansari & Gupta, 1999), it is important to estimate these correlations and elasticities using a basket-level model to eliminate the effects of misspecification bias. This is the first objective of this study, and we summarize it below:

> *Objective 1:* We estimate a basket-level model of household purchase incidence decisions to obtain estimates of pair-wise correlations across all product categories within the household's shopping basket and estimates of marketing mix elasticities in each product category.

Cross-category correlations are of interest to retailers seeking to maximize store profits by jointly coordinating marketing activities across product categories within the store. Cross-category correlations are also of interest to database marketers interested in undertaking cross-selling initiatives across product categories (Berry & Linoff, 1997). A complete basket-level model of household purchase incidence decisions, as proposed in this study, has not been estimated thus far in the marketing literature. We estimate our basket-level model using scanner panel data, which tracks the purchases of a fixed number of households across twelve different product categories in the store over time.

While using scanner panel data, it is important to investigate how sensitive the estimated cross-category correlations are to the panel structure of the data. In other words, one must assess the impact of (ignoring or accommodating) unobserved heterogeneity across households on the estimated cross-category correlations. To the extent that cross-category correlations may proxy for the effects of unobserved heterogeneity if the latter is ignored, it is possible that cross-category correlations may be overstated (and hence "spurious") in the absence of unobserved heterogeneity. Also, the estimated marketing mix elasticities in each product category may be sensitive to the inclusion of unobserved heterogeneity across households. Explicitly investigating this issue is the second objective of this study, and we summarize it below:

Objective 2: We estimate the basket-level model of household purchase incidence decisions both with and without accommodating the effects of unobserved heterogeneity across households in order to investigate the consequences of ignoring unobserved heterogeneity on the estimated cross-category correlations and households' responsiveness to marketing variables in each product category.

Disentangling cross-category correlations from unobserved heterogeneity is important to retailers since the two phenomena imply different marketing strategies. For example, if cross-category correlations are observed to be simply proxies for unaccounted-for heterogeneity across households, the marketer could develop marketing programs separately for each product category taking into account the estimated heterogeneity distribution. In such a case, separately maximizing the profits from each category is tantamount to maximizing overall store profits.

To summarize, we propose a basket-level model of household purchase incidence decisions and estimate the proposed model using scanner panel data on household purchases across twelve product categories. The proposed model has a multivariate probit panel structure and is used to estimate pair-wise correlations in households' random utilities across the twelve product categories. We employ an extension of a recently developed Bayesian method (Albert & Chib, 1993; Chib & Greenberg, 1998) to estimate model parameters. Our main findings are that either ignoring or incompletely accounting for cross-category correlations within household shopping baskets overestimates the effectiveness of marketing variables in driving purchase incidence decisions. We also find that ignoring unobserved heterogeneity across households overstates cross-category correlations and understates the effectiveness of marketing variables. The rest of the paper is organized as follows. In the next section we propose the multivariate probit panel model and discuss estimation issues. In Section 4, we provide details of the Markov Chain Monte Carlo sampling scheme. In Section 5, we give a detailed description of the data. In Section 6, we present our empirical results. We conclude with a summary and directions for future research in Section 7.

3. MODEL AND ESTIMATION

Notation

Suppose we observe binary responses of H households in J product categories over time. We refer to this collection of responses as $\{y_{htj} \in (0, 1)$: $h = 1, \ldots, H$; $t = 1, \ldots, T_h$; $j = 1, \ldots, J\}$ where subscripts h, t and j refer to household, shopping occasion and product category respectively. We define $y_{ht}=(y_{ht1}, y_{ht2}, \ldots, y_{htJ})'$, $y_h=(y'_{h1}, y'_{h2}, \ldots, y'_{hT_h})'$ and $y=(y'_1, y'_2, \ldots, y'_H)'$. Note that y_{htj} is a scalar, y_{ht} is a J-dimensional vector, y_h is a $J*T_h$-dimensional vector and y is a $\Sigma_h J*T_h$-dimensional vector.

We also observe values of k marketing variables for each product category at each shopping occasion for each household. We refer to this collection of k-dimensional covariate vectors as $\{X_{htj}: h=1, \ldots, H; t=1, \ldots, T_h; j=1, \ldots, J\}$. We define X_{ht} as

$$X_{ht}=\begin{pmatrix} X'_{ht1} & 0 & \ldots & 0 \\ 0 & X'_{ht2} & \ldots & 0 \\ 0 & 0 & \ldots & 0 \\ 0 & 0 & \ldots & X'_{htJ} \end{pmatrix}, \tag{1}$$

and define $X_h=(X'_{h1}, X'_{h2}, \ldots, X'_{hT_h})'$ and $X=(X'_1, X'_2, \ldots, X'_H)'$. Note that X_{ht} is a $(J)*(k*J)$-dimensional matrix, X_h is a $(J*Th)*(k*J)$-dimensional matrix and X is a $(\Sigma_h J*T_h)*(k*J)$-dimensional matrix.

We assume that y_{htj} not only depends on X_{htj} but also is correlated with y_{htk} (for $k \neq j$). In other words, a household's response in a product category depends both on category-specific marketing variables and on the household's responses in other product categories. This is a multivariate choice problem for the household. Previous work has either completely ignored dependencies across y'_{htj}s, thereby assuming univariate choice problems for the household for each product category (Chiang, 1991; Chintagunta, 1993), or accounted for dependencies across a limited number of obviously related product categories (Chintagunta & Haldar, 1998; Manchanda, Ansari & Gupta, 1999). In our framework we pose the multivariate choice problem in the context of the household's shopping basket, and therefore in its fullest generality. Next we present the model that explains the observed response vector y.

Multivariate Probit Model with Unobserved Heterogeneity

Let household h's latent utility at shopping occasion t for product category j be given by

$$Z_{htj}=X'_{htj}\beta_j+b_h+c_{hj}+\varepsilon_{htj}, \tag{2}$$

where X_{htj} is a k-dimensional vector of marketing variables pertaining to product category j facing household h at shopping occasion t, β_j is the corresponding k-dimensional parameter vector $(\beta_{j1}, \beta_{j2}, \ldots, \beta_{jk})$, b_h represents a household-specific random effect that is distributed $N(0, d)$, c_{hj} represents a household/category-specific random effect such that $c_h = (c_{h1}, c_{h2}, \ldots, c_{hJ})'$ is distributed $N_J(0, C)$, and ε_{htj} is a random component such that $\varepsilon_{ht} = (\varepsilon_{ht1}, \ldots, \varepsilon_{htJ})'$ is distributed $N_J(0, \Sigma)$, where Σ is a $J*J$ covariance matrix given by

$$\Sigma = \begin{pmatrix} 1 & \sigma_{12} & \ldots & \sigma_{1J} \\ & 1 & \ldots & \sigma_{2J} \\ & & \ldots & \sigma_{J-1,J} \\ & & & 1 \end{pmatrix}. \tag{3}$$

This covariance matrix is in correlation form for identifiability reasons and contains $p = J*(J-1)/2$ free parameters (see Chib & Greenberg, 1998 for details) given by $\sigma \equiv (\sigma_{12}, \sigma_{13}, \ldots, \sigma_{J-1,J})$.

It is also helpful to rewrite the model in (2) for all J categories as

$$Z_{ht} = X_{ht}\beta + i_J b_h + I_J c_h + \varepsilon_{ht}, \tag{4}$$

where $Z_{ht} = (Z_{ht1}, \ldots, Z_{htJ})'$, X_{ht} is the $(J)*(k*J)$-dimensional matrix of marketing variables facing the household at shopping occasion t (as given by Eq. (1)), β is the corresponding $k*J$-dimensional parameter vector $(\beta_1, \beta_2, \ldots, \beta_J)$ where $\beta_j = (\beta_{j1}, \beta_{j2}, \ldots, \beta_{jk})$, i_j is a J-dimensional vector of ones, I_J is a $J*J$ identity matrix, b_h is a household-specific (scalar) random effect that is distributed $N(0, d)$, and c_h is a J-dimensional household-specific random effect vector that is distributed $N_J(0, C)$. Observed responses y_{htj} are determined by the unobserved latent variables Z_{htj} as:

$$y_{htj} = I[Z_{htj} > 0], \tag{5}$$

where I is the indicator function. This completes the specification of our model. The total number of parameters in the proposed model is equal to $J*[k + (J-1)/2 + 1]$ (i.e. $k*J$ covariate coefficients, plus $J*(J-1)/2$ correlation coefficients, plus $J*1$ random effects parameters).[2] If the number of product categories J is small (say, 2–4) we obtain the cross-category model of Manchanda, Ansari and Gupta (1999). If the random effects are restricted to be the same across product categories, i.e. c_h is ignored, we obtain a restricted version of our proposed model that assumes the unobserved heterogeneity distribution to be common across product categories. If the effects of unobserved heterogeneity are ignored altogether (i.e. b_h and c_h are ignored), we obtain a cross-sectional version (as opposed to a panel version) of our proposed multivariate probit model (as in Chib & Greenberg, 1998). If correlations across product categories are ignored in the common random effects model, i.e. $\Sigma = I$ (a diagonal matrix of ones), we obtain J independent category models with a common unobserved heterogeneity distribution. If the unobserved heterogeneity distributions are assumed to be independent across product categories, we obtain single-category heterogeneous models as in Chiang (1991), Bucklin and Lattin (1991), Chintagunta (1993) etc.

We can estimate marketing mix elasticities for each product category based on our proposed model and compare these elasticities to those obtained using a model that ignores cross-category correlations Σ. This allows us to understand the effects of ignoring cross-category correlations on measures of managerial relevance such as price elasticities (our research objective no. 1). We can also compare the correlation matrix Σ estimated using our proposed model with that estimated using a restricted version of the model that ignores household-specific random effects (i.e. b_h and c_h). This allows us to understand the effects of ignoring unobserved heterogeneity across households on the estimated cross-category correlations (our research objective no. 2).

Given J product categories and T_h observations for a given household h, likelihood-based estimation of our proposed model requires the computation of the likelihood contribution

$$\begin{aligned}\Pr(y_h \mid \beta, \sigma, d, C) &= \int \left[\prod_{t=1}^{T_h} \int_{B_{hJt}} \int_{B_{hJ-1,t}} \cdots \int_{B_{h1t}} \phi_J(Z_{ht} \mid X_{ht}\beta + i_J b_h + I_J c_h, \Sigma) dZ_{ht} \right) \Bigg] \\ &\quad \cdot \phi(b_h \mid 0, d)\phi(c_h \mid 0, C) db_h dc_h, \end{aligned} \tag{6}$$

for each household $h = 1, \ldots, H$, where $\phi_J(\,.\mid \mu, \Sigma)$ is the density of a J-variate normal distribution with mean μ and covariance matrix Σ, B_{hjt} is the interval $(0, \infty)$ if $y_{hjt} = 1$ and the interval $(-\infty, 0)$ if $y_{hjt} = 0$. This likelihood contribution is quite difficult to compute even using simulation techniques. Given the computational intractability of likelihood-based estimation, we adopt a simulation-based Bayesian approach to estimate model parameters.

Bayesian Approach to Model Estimation

Given the response vector y, the matrix of covariates X, and a prior density on model parameters given by $\pi(\beta, \sigma, d, C)$, Bayes rule yields

$$\pi(\beta, \sigma, d, C \mid y) \propto \pi(\beta, \sigma, d, C) * \Pr(y \mid \beta, \sigma, d, C), \tag{7}$$

where

$$\Pr(y \mid \beta, \sigma, d, C) = \left(\prod_{h=1}^{H} \Pr(y_h \mid \beta, \sigma, d, C) \right) * U[\sigma \in Q], \tag{8}$$

and $\Pr(y_h \mid \beta, \sigma, d, C)$ is given by Eq. (6) and Q is a convex solid body in the hypercube $[-1, 1]^p$ that leads to a proper correlation matrix. This form of the posterior density is not particularly useful for Bayesian estimation since it involves the evaluation of the complicated likelihood function (just as in likelihood-based estimation). Instead of attempting to directly evaluate the joint posterior density we invoke the data augmentation framework of Albert and Chib (1993) and Chib and Greenberg (1998). This framework is based on taking a sampling-based approach, in conjunction with Markov Chain Monte Carlo (MCMC) techniques (Tanner & Wong, 1987; Gelfand & Smith, 1990; Tierney, 1994; Chib & Greenberg, 1995), based on the conditional distributions given by

$$Z_{ht} \mid y_h, \beta, \sigma, d, C; \; t = 1, \ldots, T_h; \; h = 1, \ldots, H$$

$$\beta \mid y_h, Z_h, \sigma, d, C$$

$$b_h \mid y_h, Z_h, \beta, \sigma, C; \; h = 1, \ldots, H$$

$$c_h \mid y_h, Z_h, \beta, \sigma, d; \; h = 1, \ldots, H$$

$$\sigma \mid y_h, Z_h, \beta, d, C$$

$$d^{-1} \mid y_h, Z_h, \beta, \sigma, C$$

$$C^{-1} \mid y_h, Z_h, \beta, \sigma, d$$

Each of these distributions (except that of σ) is of known form and can be sampled directly. Details are provided in the next section. The key simplification that data augmentation provides in our context is that it allows us to bypass the computation of the likelihood.

4. MARKOV CHAIN MONTE CARLO (MCMC) SAMPLING

Prior Distributions

For the purposes of our analysis, we assume that our prior information can be represented by the distributions

$$\beta \sim N_{k*J}(\beta_o, B_o),$$

$$\sigma \sim N_{J*(J-1)/2}(g_o, G_o) * I[\sigma \in Q],$$

$$d^{-1} \sim G(\eta_o, \psi_o),$$

$$C^{-1} \sim \text{Wish}_J(\rho_o, R_o),$$

where the hyperparameters are as follows: β_o is a $k*J$-dimensional vector of zeros, B_o is a $(k*J)*(k*J)$ diagonal-matrix, with its diagonal elements equal to 0.1 implying a variance of 10 for each component of β, g_o is a p-dimensional vector with all its elements equal to 0.5, G_o is a $p*p$ identity matrix, $\eta_o = 1$, $\chi_o = 3$, $\rho_o = \mathrm{J} + 4$, $R_o = 3*I_{J*J}$. The choice of these priors is intended to represent vague prior information.

MCMC Algorithm

We are interested in simulating from the posterior distribution of $(\{Z_h\}, \beta, \sigma, \{b_h\}, \{c_h\}, d^{-1}, C^{-1})$, where Z_h is a $J*T_h$-dimensional vector given by $(Z'_{h1} Z'_{h2} \ldots Z'_{hTh})'$. While it is difficult to sample from the joint posterior, it is possible to simulate from the conditional distributions $f(Z_h \mid \beta, \sigma, d^{-1}, C^{-1})$, $\pi(\beta \mid \{Z_h\}, \sigma, d^{-1}, C^{-1})$, $\pi(b_h, c_h \mid \{Z_h\}, \beta, \sigma, C^{-1})$, $\pi(\sigma \mid \{Z_h\}, \beta, \{b_h\}, \{c_h\}, d^{-1}, C^{-1})$ and $\pi(D^{-1} \mid \{Z_h\}, \beta, \sigma, \{b_h\}, \{c_h\})$. The MCMC sampling algorithm works as follows.

Step 0: Initialize β to $\beta^{(o)}$, σ to $\sigma^{(o)}$, set $g = 1$
Step 1: Draw $Z_h^{(g)}$ from $f(Z_h \mid y, \beta^{(g-1)}, \sigma^{(g-1)}, d^{-1(g-1)}, C^{-1(g-1)})$, $h = 1, \ldots, H$.
Step 2: Draw $\beta^{(g)}$ from $\pi(\beta \mid y, \{Z_h^{(g)}\}, \sigma^{(g-1)}, d^{-1(g-1)}, C^{-1(g-1)})$.
Step 3: Draw $b_h^{(g)}$ from $\pi(b_h \mid y, \{Z_h^{(g)}\}, \beta^{(g)}, c_h^{(g-1)}, \sigma^{(g-1)}, d^{-1(g-1)}, C^{-1(g-1)})$, $h = 1, \ldots, H$.
Step 4: Draw $c_h^{(g)}$ from $\pi(c_h \mid y, \{Z_h^{(g)}\}, \beta^{(g)}, b_h^{(g)}, \sigma^{(g-1)}, d^{-1(g-1)}, C^{-1(g-1)})$, $h = 1, \ldots, H$.
Step 5: Draw $\sigma^{(g)}$ from $\pi(\sigma \mid y, \{Z_h^{(g)}\}, \beta^{(g)}, \{b_h^{(g)}\}, \{c_h^{(g)}\}, d^{-1(g-1)}, C^{-1(g-1)})$.
Step 6: Draw $d^{-1(g)}$ from $\pi(d^{-1} \mid y, \{Z_h^{(g)}\}, \beta^{(g)}, \sigma^{(g)}, \{b_h^{(g)}\}, \{c_h^{(g)}\}, C^{-1(g-1)})$.
Step 7: Draw $C^{-1(g)}$ from $\pi(C^{-1} \mid y, \{Z_h^{(g)}\}, \beta^{(g)}, \sigma^{(g)}, \{b_h^{(g)}\}, \{c_h^{(g)}\}, d^{-1(g-1)})$.
Step 8: $g = g + 1$.
Step 9: Go to step 1.

The above cycle of seven steps is repeated a large number of times (in our example, the entire simulation is run for 10,000 cycles). From the theory of MCMC simulations, it follows that the draws on $\theta = (\{Z_h\}, \beta, \{b_h\}, \{c_h\}, \sigma, d^{-1}, C^{-1})$, beyond a burn-in period of say 500 iterations, may be taken as draws from the posterior distribution of θ. Therefore, on the basis of the simulated sample, we are able to obtain point and interval estimates of the parameters and other summaries of the posterior distribution. Next, we provide the form of each of the five conditional distributions given in steps 1–7.

1. $Z_h \mid y, \beta, \sigma, d^{-1}, C^{-1} \propto \mathrm{N}_{J*Th}(Z_h \mid X_h\beta, i_{JTh}d + (i_{JTh} \otimes I_J)C(i_{JTh} \otimes I_J)' + I_{Th} \otimes \Sigma) * \Pi_t \Pi_j \{\mathrm{I}(Z_{htj} > 0) * \mathrm{I}(y_{htj} = 1) + \mathrm{I}(Z_{htj} \leq 0) * \mathrm{I}(y = 0)\}$, where i_{JTh} is a $J*T_h$-

dimensional vector of ones, I_J is a $J*J$ identity matrix, and I_{JTh} is a JT_h*JT_h identity matrix. This is a truncated multivariate normal distribution. This distribution is sampled through a Gibbs cycle (see Geweke, 1991). This representation of the conditional posterior of Z_h follows from Albert and Chib (1993).

2. $\beta | y, \{Z_{ht}\}, \sigma, d^{-1}, C^{-1} \sim N_{k*J}(\beta | \hat{\beta}, B)$, where $B = (B_o + X'_h(I_{Th} \otimes \Sigma)^{-1} X_h)^{-1}$, $\hat{\beta} = B(B_o^{-1}\beta_o + \Sigma_h X'_h(I_{Th} \otimes \Sigma)^{-1} Z_h)$.
3. $b_h | y, \{Z_{ht}\}, \beta, \sigma, d^{-1}, C^{-1}, c_h \sim N_J(b_h | \hat{b}_h, B_h)$, where $B_h = ((dI_J)^{-1} + \Sigma_t I'_J I_J)^{-1}$, $\hat{b}_h = B_h(\Sigma_t I'_J(Z_{ht} - X_{ht}\beta - I_J c_h))$, $h = 1, \ldots, H$.
4. $c_h | y, \{Z_{ht}\}, \beta, \sigma, d^{-1}, C^{-1}, b_h \sim N_J(c_h | \hat{b}_h, \hat{B}_c)$, where $\hat{B}_c = (C^{-1} + \Sigma_t I'_J I_J)^{-1}$, $\hat{b}_c = \hat{B}_c(\Sigma_t I'_J(Z_{ht} - X_{ht}\beta - i_J b_h))$, $h = 1, \ldots, H$.
5. $\sigma | y, \{Z_{ht}\}, \beta, \{b_h\}, \{c_h\}, d^{-1}, C^{-1} \sim N_{J*(J-1)/2}(g_o, G_o) * I[\sigma \in Q] * \Pi_h \Pi_t N_J(Z_{ht} | X_{ht}\beta + i_J b_h + I_J c_h, \Sigma)$. We use the Metropolis-Hastings algorithm to sample from this non-standard distribution (details given in the next subsection), following Chib and Greenberg (1998).
6. $d^{-1} | y, \beta, \{b_h\}, \{c_h\}, \sigma \sim IG(\eta_o + H, \chi)$, where $\chi = (\chi_o^{-1} + \Sigma_h b_h b'_h)^{-1}$.
7. $C^{-1} | y, \beta, \{b_h\}, \{c_h\}, \sigma \sim W_J(\rho_o + H, R)$, where $R = (R_o^{-1} + \Sigma_h c_h c'_h)^{-1}$.

Metropolis-Hastings (M-H) Algorithm

The only distribution in the set above that cannot be sampled directly is the distribution of σ, i.e. $\pi(\sigma | y, \{Z_{ht}\}, \beta, \{b_h\}, \{c_h\}, d^{-1}, C^{-1})$. To sample this distribution we use the M-H algorithm (see Chib & Greenberg, 1995 for a detailed exposition). Suppose $q(\sigma | \sigma', y, \{Z_{ht}\}, \beta, \{b_h\}, \{c_h\}, D^{-1})$ is a candidate generating density. Then to draw σ we proceed as follows.

Step 1: Sample a proposal value σ' given σ from $q(\sigma' | y, \{Z_{ht}\}, \beta, \{b_h\}, \{c_h\}, D^{-1})$.

Step 2: Move to σ' with probability $\alpha(\sigma, \sigma')$ and stay at σ with probability $1 - \alpha(\sigma, \sigma')$, where

$\alpha(\sigma, \sigma') =$

$$\min\left\{\frac{\pi(\sigma' | y, Z_{ht}, \beta, \{b_h\}, \{c_h\}, D^{-1})}{\pi(\sigma | y, Z_{ht}, \beta, \{b_h\}, \{c_h\}, D^{-1})} * \frac{q(\sigma | \sigma', y, Z_{ht}, \beta, \{b_h\}, \{c_h\}, D^{-1})}{q(\sigma' | \sigma, y, Z_{ht}, \beta, \{b_h\}, \{c_h\}, D^{-1})}\right\},$$

We use the *tailored chain* as our choice of candidate generating density, as in Chib and Greenberg (1998). It is specified as

$$\sigma' = \mu + g,$$

where μ is a p-dimensional vector, taken to be the mode of log $\pi(\sigma | y, \{Z_{ht}\}, \beta, \{b_h\}, \{c_h\}, d^{-1}, C^{-1})$ and $g \sim MVt(0, \tau V, \nu)$, where V is the negative of the second derivative of log $\pi(\sigma | y, \{Z_{ht}\}, \beta, \{b_h\}, \{c_h\}, d^{-1}, C^{-1})$ evaluated at the mode. This approach leads to a well mixing Markov chain.

5. DESCRIPTION OF DATA

We employ IRI's scanner panel database on household purchases in twenty-five product categories in a metropolitan market in a large U.S. city. For our analysis, we pick twelve product categories: bacon, butter, coffee, cola, crackers, detergent, hot dogs, ice cream, non-cola beverages, sugar, toilet tissue and paper towels. These product categories[3] have been identified in the literature as being representative of the household's "shopping basket" (see Bell & Lattin, 1998). The dataset covers a period of two years from June 1991 to June 1993 and contains shopping visit information on 494 panelists across four different stores in an urban market. For each product category, the dataset contains information on marketing variables – price, in-store displays, and newspaper feature advertisements – at the SKU-level for each store/week.

Choosing households that bought at the two largest stores in the market (that collectively account for 90% of all shopping visits in the database) yields 488 households. From these households, we pick a random sample of 300 households making a total of 39,276 shopping visits at the two largest stores. This is done to keep the size of the dataset manageable. For those shopping visits when a household visits the store but does not purchase a particular product category, we compute marketing variables as share-weighted average values across all SKUs in the product category, where shares are household-specific and computed using the observed purchases of the household over the study period. Computing marketing variables using such share-weighting has precedence in the empirical marketing literature on category purchase incidence[4] (see, for example, Manchanda, Ansari & Gupta, 1999). Descriptive statistics pertaining to the marketing variables are provided in Table 1.

From Table 1 we can see that average display and feature activity is higher for purchase visits than for non-purchase visits, as expected, for all product categories. In terms of the magnitude of the difference in display and feature activity between purchase and non-purchase visits, the largest magnitude is observed for toilet tissue, suggesting that store merchandising activities strongly influence household purchase incidence for this product category. The smallest magnitudes are observed for ice-cream and non-cola beverages for display and feature respectively. Average prices are lower for purchase visits than for non-purchase visits, as expected, for ten out of the twelve product categories. By and large, these descriptive statistics are consistent with the economic notions of positive own-advertising elasticities, negative own-price

Table 1. Descriptive Statistics on Marketing Variables
Number of households = 300, Number of shopping visits = 39,276.

A. Purchase visits

Product	**Price ($/RP)**	**Display**	**Feature**	**No. of Purchases**
Bacon	1.7915	0.2078	0.5338	2473
Butter	1.0425	0.1910	0.3079	5787
Coffee	1.9107	0.3174	0.3439	3022
Cola	0.6033	0.3999	0.4749	5099
Crackers	2.9236	0.2093	0.1280	4214
Detergent	0.8991	0.3550	0.2840	3159
Hot dogs	2.0753	0.1564	0.3832	3847
Ice cream	0.7196	0.0019	0.3964	4334
Non-cola	0.6654	0.1963	0.1340	5922
Sugar	0.4565	0.3681	0.3820	2275
Tissue	0.3041	0.4084	0.4457	5534
Towels	0.7386	0.3561	0.3544	4482

B. Non-purchase visits

Product	**Price ($/RP)**	**Display**	**Feature**	**No. of Visits**
Bacon	2.2949	0.0739	0.2333	36,803
Butter	1.1089	0.0686	0.1169	33,489
Coffee	2.0284	0.1074	0.0998	36,254
Cola	0.7080	0.1392	0.2306	34,177
Crackers	2.6717	0.1003	0.0569	35,062
Detergent	1.1150	0.0937	0.0547	36,117
Hot dogs	2.4145	0.0461	0.1612	35,429
Ice cream	0.8042	0.0008	0.1585	34,942
Non-cola	0.6736	0.1086	0.0779	33,354
Sugar	0.4456	0.1193	0.1197	37,001
Tissue	0.3369	0.1236	0.1345	33,742
Towels	0.8081	0.1159	0.1060	34,794

elasticities etc. From the last column of Table 1, we can see that the most frequently purchased product category is non-cola beverages (with butter coming second), while the most infrequently purchased product category is sugar (with bacon coming second).

In Table 2a, we report, in matrix form, the purchase frequencies for each product category along the diagonal and pair-wise purchase frequencies for each pair of product categories (i.e. the number of times each pair of product

Table 2.

A: Descriptive Statistics – Joint Purchase Frequencies

	Bacon	Butter	Coffee	Cola	Crackers	Deterg.	Hot dogs	Ice cream	Non-cola	Sugar	Tissue	Towels
Bacon	**2473**											
Butter	710	**5787**										
Coffee	324	799	**3022**									
Cola	502	1316	620	**5099**								
Crackers	428	1198	597	992	**4214**							
Deterg.	382	879	488	844	752	**3159**						
Hot dogs	653	1091	468	927	737	608	**3847**					
Ice cream	415	1046	542	817	797	581	702	**4334**				
Non-cola	624	1341	618	1726	1104	822	1018	1035	**5922**			
Sugar	338	772	328	478	431	359	453	400	573	**2275**		
Tissue	719	1694	858	1389	1180	1127	1051	989	1467	661	**5534**	
Towels	490	1308	751	1220	1020	919	791	823	1221	552	1897	**4482**

B: Descriptive Statistics – Bivariate Correlations

	Bacon	Butter	Coffee	Cola	Crackers	Deterg.	Hot dogs	Ice cream	Non-cola	Sugar	Tissue	Towels
Bacon	**1**											
Butter	0.1022	**1**										
Coffee	0.0526	0.0953	**1**									
Cola	0.0564	0.1207	0.0647	**1**								
Crackers	0.0551	0.1339	0.0842	0.1089	**1**							
Deterg.	0.0706	0.1092	0.0860	0.1208	0.1249	**1**						
Hot dogs	0.1449	0.1267	0.0553	0.1090	0.0897	0.0940	**1**					
Ice cream	0.0475	0.0934	0.0636	0.0615	0.0872	0.0694	0.0759	**1**				
Non-cola	0.0736	0.0940	0.0433	0.2026	0.1077	0.0904	0.1048	0.0866	**1**			
Sugar	0.0874	0.1343	0.0626	0.0592	0.0658	0.0705	0.0844	0.0518	0.0701	**1**		
Tissue	0.1116	0.1814	0.1187	0.1460	0.1386	0.1835	0.1253	0.0884	0.1294	0.1066	**1**	
Towels	0.0685	0.1463	0.1220	0.1520	0.1395	0.1644	0.0948	0.0839	0.1220	0.1002	0.2913	**1**

categories is purchased together) along the off-diagonal. For example, bacon is purchased on 2473 shopping visits, of which 710 are associated with the joint purchase of butter. This means that 28.7% of all bacon purchases are associated with joint purchase of butter. We report bivariate rank correlations, based on these purchase frequencies, in the lower half (i.e. below the main diagonal) of Table 2b. Cross-category correlations are fairly evident, with high magnitudes observed for two pairs: tissue and towels (0.2913), non-cola and cola beverages (0.2026).

All the observed correlations in Table 2b are positive. The reason for this is the large number of "zeros" that characterizes the vector of purchase outcomes for each product category. For example, among the 39,276 store visit observations in the dataset, only 5922 resulted in the purchase of non-cola beverages, 5534 resulted in the purchase of tissue, etc. This means that product categories appear to be complements for no reason other than the fact neither was purchased on a large number of purchase occasions. One way to "correct" for this is to recompute bivariate correlations for each pair after ignoring observations that resulted in a purchase of neither (let us call these "zero observations"). But this creates a problem of the opposite kind, i.e. all pairs of product categories appear to be substitutes on account of our ignoring a large number of outcomes when neither is purchased. However, the amount of distortion observed in the bivariate correlation for a given pair of product categories when its zero observations are ignored, is almost identical to the distortion observed for any other pair of product categories when their zero observations are ignored. This means that comparing bivariate correlations across pairs of product categories is meaningful, regardless of how we compute the correlations. For example, toilet tissue and towels have a much higher bivariate correlation than bacon and coffee regardless of whether or not we ignore each pair's zero observations. Armed with these preliminary findings, we next estimate our proposed econometric model on the basket data in order to estimate cross-category relationships after accommodating the effects of covariates, panel structure of the data etc. While estimating the proposed model, we include the following variables in the household-specific vector X_{ht} (see Eq. 2) for each of the twelve product categories in the shopping basket.

1. Price
2. Feature
3. Display
4. Inventory

Price is a continuous variable, operationalized in dollars per ounce. Feature and display are indicator variables, that take the value 1 if the product is on feature or display respectively, and 0 otherwise. Inventory is a continuous variable (measured in ounces per week), which is computed using the household's product consumption rate which, in turn, is computed by dividing the total product quantity purchased by the household over the study period by the number of weeks in the data. For the first week in the data, each household is assumed to have enough inventory for that week, i.e. the inventory variable for a household at $t = 1$ is assumed to be the household's weekly product consumption rate. We incorporate random effects in the intercept terms for each product category.

6. EMPIRICAL RESULTS

We estimate the proposed basket-level model of purchase incidence decisions as well as five benchmark models, as shown below, in order to investigate the consequences of ignoring either cross-category correlations or unobserved heterogeneity across households.

Model 1: Multivariate Probit – Full twelve categories
Model 2: Multivariate Probit with unobserved heterogeneity restricted to be common across categories – Full twelve categories
Model 3: Multivariate Probit – Four categories only
Model 4: Multivariate Probit – Two categories only
Model 5: Independent Univariate Probits
Model 6: Multivariate Probit without unobserved heterogeneity

Comparing model 1 vs. model 2 allows one to investigate the consequences of restricting the unobserved heterogeneity distribution to be the same across product categories. For models 3 and 4, we retain the assumption of common unobserved heterogeneity distribution across product categories (as in model 2). Comparing model 2 vs. models 3 and 4 will demonstrate the consequences of modeling households' purchase incidence decisions only across subsets of the twelve product categories. For model 5, we assume the unobserved heterogeneity distribution to be different across product categories (as in model 1). Comparing model 1 vs. model 5 will demonstrate the consequences of modeling purchase incidence decisions jointly as opposed to separately across product categories. Comparing models 1 or 2 vs. model 6 will demonstrate the consequences of ignoring unobserved heterogeneity across households in a multivariate probit model.

First we look at the estimated inter-category correlation matrix based on the proposed multivariate probit model, allowing the unobserved heterogeneity distribution to be different across product categories (i.e. model 1). This is summarized in Table 3. The lower triangle reports the posterior means, while the upper triangle reports the 95% posterior credibility intervals (symmetric about the posterior mean). The off-diagonal terms in this table indicate that inter-category correlations are non-zero in general, with the correlations being quite large for specific pairs of product categories. For example, the estimated correlation in purchase incidence outcomes between cola and non-cola beverages has a mean of 0.4216 and a credibility interval of (0.40, 0.45). This indicates that households, rather than viewing cola and non-cola beverages as consumption substitutes, buy them together for complementary consumption needs, i.e. to maintain variety in their "beverage pantry." The estimated correlation is also large for hot dogs and bacon (0.3812), another possible consequence of the household's need for variety in the kitchen, this time among the meat products in their refrigerator. A third pair of product categories for which the estimated correlation is high is tissue and detergents (0.3744). This finding is especially interesting since there is little opportunity for a *sheer coincidence* effect, i.e. the two product categories frequently co-occurring in the household's shopping basket on account of having short inter-purchase cycles. In fact, inter-purchase times in these product categories are much larger, on average, than for other product categories in the data. One possible explanation for the large value of the estimated correlation is that since detergents and tissue are typically shelved close to each other in the grocery store, frequently in the same aisle, households have a propensity to pick up both products at the same time. One managerial implication of this "shelf effect" phenomenon is that the retailer may improve store profitability by shelving high-margin product categories close to products with short inter-purchase cycles so that every time a consumer picks up the latter off store shelves, she faces an opportunity to pick up the nearby high-margin product as well.

In Table 4, we report the estimated cross-category correlations using model 2 that assumes the unobserved heterogeneity distribution to be common across product categories. A comparison of Tables 3 and 4 indicates that cross-category correlations are, by and large, understated in Table 4 (i.e. model 2). To the extent that the common unobserved heterogeneity distribution across product categories captures correlations in households' purchase outcomes across categories,[5] one would indeed expect any remaining cross-category correlations in purchase outcomes to decrease after accounting for such unobserved heterogeneity.

In Table 5, we report the estimated cross-category correlations using model 3 that looks at four product categories at a time (as in Manchanda, Ansari & Gupta, 1999). A comparison of Tables 4 and 5 indicates that ignoring the

Table 3. Estimated Pair-Wise Correlations across Product Categories – MVP on 12 Categories with Different Unobserved Heterogeneity across Categories (Model 1).[10]

	Bacon	Butter	Coffee	Cola	Crackers	Deterg.	Hot dogs	Ice cream	Non-cola	Sugar	Tissue	Towels
Bacon	1	0.19, 0.24	0.15, 0.24	0.15, 0.25	0.12, 0.21	0.13, 0.24	0.35, 0.41	0.10, 0.19	0.15, 0.22	0.19, 0.29	0.31, 0.38	0.20, 0.30
Butter	0.2177	1	0.18, 0.23	0.20, 0.26	0.23, 0.29	0.19, 0.26	0.27, 0.33	0.18, 0.24	0.15, 0.21	0.31, 0.38	0.29, 0.35	0.24, 0.31
Coffee	0.1936	0.2068	1	0.14, 0.19	0.18, 0.26	0.18, 0.27	0.19, 0.26	0.19, 0.26	0.12, 0.20	0.22, 0.32	0.27, 0.34	0.23, 0.32
Cola	0.2067	0.2291	0.1640	1	0.16, 0.20	0.17, 0.26	0.19, 0.25	0.15, 0.24	0.40, 0.45	0.10, 0.21	0.22, 0.30	0.24, 0.31
Crackers	0.1698	0.2612	0.2183	0.1820	1	0.16, 0.20	0.17, 0.23	0.16, 0.23	0.20, 0.27	0.16, 0.24	0.23, 0.30	0.24, 0.31
Deterg.	0.1938	0.2223	0.2246	0.2144	0.1791	1	0.15, 0.22	0.11, 0.21	0.16, 0.23	0.24, 0.31	0.34, 0.41	0.30, 0.37
Hot dogs	0.3812	0.2964	0.2226	0.2194	0.1992	0.1872	1	0.12, 0.17	0.19, 0.26	0.22, 0.32	0.25, 0.32	0.20, 0.29
Ice cream	0.1414	0.2089	0.2267	0.1930	0.1907	0.1609	0.1447	1	0.12, 0.16	0.10, 0.20	0.15, 0.23	0.16, 0.25
Non-cola	0.1816	0.1793	0.1651	0.4216	0.2355	0.1931	0.2253	0.1419	1	0.12, 0.17	0.23, 0.29	0.20, 0.26
Sugar	0.2362	0.3423	0.2709	0.1512	0.1999	0.2700	0.2667	0.1528	0.1476	1	0.22, 0.27	0.28, 0.37
Tissue	0.3449	0.3199	0.3064	0.2576	0.2654	0.3744	0.2900	0.1870	0.2592	0.2471	1	0.27, 0.31
Towels	0.2532	0.2742	0.2779	0.2756	0.2695	0.3351	0.2462	0.2004	0.2326	0.3214	0.2915	1

Table 4. Estimated Pair-Wise Correlations across Product Categories – MVP on 12 Categories with Common Unobserved Heterogeneity across Categories (Model 2).[11]

	Bacon	Butter	Coffee	Cola	Crackers	Deterg.	Hot dogs	Ice cream	Non-cola	Sugar	Tissue	Towels
Bacon	1	0.19, 0.24	0.11, 0.20	–0.01, 0.07	0.04, 0.12	0.11, 0.20	0.35, 0.41	0.09, 0.15	0.12, 0.19	0.23, 0.34	0.20, 0.27	0.06, 0.15
Butter	0.2205	1	0.18, 0.23	0.14, 0.19	0.22, 0.26	0.17, 0.25	0.21, 0.27	0.13, 0.18	0.08, 0.13	0.28, 0.33	0.27, 0.32	0.22, 0.27
Coffee	0.1475	0.2071	1	0.03, 0.10	0.16, 0.22	0.17, 0.23	0.12, 0.19	0.14, 0.20	0.02, 0.08	0.14, 0.23	0.21, 0.28	0.24, 0.33
Cola	0.0381	0.1627	0.0665	1	0.06, 0.14	0.12, 0.19	0.08, 0.14	–0.03, 0.05	0.29, 0.33	0.00, 0.08	0.16, 0.23	0.21, 0.27
Crackers	0.0896	0.2401	0.1922	0.1020	1	0.23, 0.29	0.12, 0.17	0.14, 0.20	0.13, 0.19	0.11, 0.18	0.18, 0.24	0.21, 0.28
Deterg.	0.1519	0.2077	0.2055	0.1518	0.2550	1	0.12, 0.19	0.06, 0.13	0.09, 0.15	0.18, 0.25	0.30, 0.36	0.32, 0.37
Hot dogs	0.3819	0.2397	0.1486	0.1123	0.1408	0.1529	1	0.13, 0.19	0.14, 0.21	0.22, 0.29	0.15, 0.22	0.12, 0.18
Ice cream	0.1224	0.1536	0.1724	0.0172	0.1709	0.0910	0.1612	1	0.12, 0.17	0.08, 0.15	0.06, 0.12	0.08, 0.14
Non-cola	0.1554	0.1058	0.0538	0.3107	0.1582	0.1269	0.1758	0.1479	1	0.12, 0.18	•0.15, 0.20	0.15, 0.21
Sugar	0.2821	0.3085	0.1941	0.0379	0.1513	0.2205	0.2510	0.1145	0.1521	1	0.22, 0.28	0.18, 0.25
Tissue	0.2340	0.2953	0.2464	0.1936	0.2105	0.3344	0.1875	0.0907	0.1707	0.2490	1	0.50, 0.54
Towels	0.1129	0.2491	0.2845	0.2354	0.2442	0.3467	0.1491	0.1125	0.1727	0.2104	0.5239	1

Table 5. Estimated Pair-Wise Correlations across Product Categories – MVP on 4 Categories with Common Unobserved Heterogeneity across Categories (Model 3).[12]

	Bacon	Butter	Coffee	Cola	Crackers	Deterg.	Hot dogs	Ice cream	Non-cola	Sugar	Tissue	Towels
Bacon	1	0.16, 0.22	0.09, 0.18	−0.07, 0.01								
Butter	0.1927	1	0.14, 0.21	0.06, 0.12								
Coffee	0.1388	0.1729	1	−0.05, 0.03								
Cola	−0.0289	0.0877	−0.0077	1								
Crackers					1	0.18, 0.24	0.06, 0.13	0.06, 0.13				
Deterg.					0.2114	1	0.08, 0.16	−0.01, 0.06				
Hot dogs					0.0985	0.1197	1	0.06, 0.13				
Ice cream					0.1002	0.0296	0.0982	1				
Non-cola									1	0.07, 0.14	0.07, 0.13	0.06, 0.13
Sugar									0.1044	1	0.16, 0.24	0.14, 0.21
Tissue									0.0949	0.2036	1	0.45, 0.51
Towels									0.0954	0.1761	0.4804	1

remaining eight product categories within the shopping basket understates the estimated correlation in purchase incidence decisions across the included four product categories. In fact, for two pairs of product categories ([cola & bacon] and [cola & coffee]), the estimated correlations are negative in the four-variate probit model even though they are positive in the twelve-variate probit model. For example, the posterior mean and credibility interval of the correlation for the pair [cola & bacon], based on model 3, are –0.0289 and (–0.0683, 0.0100) respectively. The corresponding measures based on model 2 are 0.0381 and (–0.0099, 0.0735) respectively. Similarly, the posterior mean and credibility interval of the correlation for the pair [cola & coffee], based on model 3, are –0.0077 and (–0.0446, 0.0264) respectively. The corresponding measures based on model 2 are 0.0665 and (0.0345, 0.1031) respectively. This indicates that if one were to use model 3, instead of model 2, one may falsely conclude, for example, that cola and coffee substitute for each other within the household's shopping basket when, in fact, they do not!

In Table 6, we report the estimated cross-category correlations using model 4 – that looks at pairs of product categories only (as in Chintagunta & Haldar, 1998) – for nine different pairs of product categories. A comparison of Tables 4 and 6 indicates that ignoring the remaining ten product categories within the shopping basket understates the estimated correlation in purchase incidence decisions for each pair of product categories. In fact, for three pairs of product categories – [cola & sugar], [cola & coffee], [cola & crackers] – the estimated correlations are negative in the bivariate probit model even though they are positive in the twelve-variate probit model. For example, the posterior mean and credibility interval of the correlation for the pair [cola & coffee], based on model 4, are –0.0733 and (–0.1100, –0.0367) respectively. The corresponding measures based on model 2 are 0.0665 and (0.0345, 0.1031) respectively. This indicates that if one were to use model 4, instead of model 2, one may falsely that cola and coffee, substitute each other within the household's shopping basket when, in fact, they do not! We summarize this finding below.

> *Empirical Finding 1:* A limited operationalization of the multivariate probit model with panel structure, using a subset of the full set of product categories within the household's shopping basket (as in Chintagunta & Haldar, 1998; Manchanda, Ansari & Gupta, 1999), leads one to underestimate correlations in households' purchase incidence decisions across product categories. The estimated correlations even change signs (from positive to negative) in a few cases.

In Table 7, we report the estimated cross-category correlations using model 6 that ignores unobserved heterogeneity across households, i.e. a cross-sectional

Table 6. Estimated Pair-Wise Correlations across Product Categories – MVP on 2 Categories with Common Unobserved Heterogeneity across Categories (Model 4).

Model/ Categories	Bacon & Hot dogs	Butter & Sugar	Cola & Sugar	Cola & Non-cola	Detergent & Tissue	Detergent & Towels	Tissue & Towels	Cola & Coffee	Cola & Crackers
Mean	0.2498	0.2228	–0.1172	0.1828	0.2166	0.2169	0.4050	–0.0733	–0.0190
C.I.[13]	0.20, 0.29	0.18, 0.26	–0.18, –0.06	0.15, 0.21	0.17, 0.26	0.17, 0.26	0.37, 0.45	–0.11, –0.04	–0.05, 0.01

Table 7. Estimated Pair-Wise Correlations across Product Categories – MVP on 12 Categories without Unobserved Heterogeneity (Model 6).[14]

	Bacon	Butter	Coffee	Cola	Crackers	Deterg.	Hot dogs	Ice cream	Non-cola	Sugar	Tissue	Towels
Bacon	1	0.31, 0.36	0.20, 0.28	0.17, 0.24	0.20, 0.26	0.26, 0.33	0.43, 0.49	0.16, 0.24	0.24, 0.31	0.31, 0.39	0.35, 0.40	0.24, 0.31
Butter	0.3360	1	0.30, 0.36	0.31, 0.36	0.35, 0.39	0.34, 0.39	0.33, 0.38	0.24, 0.29	0.23, 0.28	0.37, 0.44	0.42, 0.47	0.39, 0.43
Coffee	0.2372	0.3302	1	0.21, 0.26	0.27, 0.33	0.30, 0.37	0.21, 0.28	0.22, 0.28	0.15, 0.20	0.22, 0.30	0.35, 0.41	0.39, 0.45
Cola	0.2055	0.3344	0.2361	1	0.26, 0.30	0.30, 0.37	0.25, 0.31	0.15, 0.21	0.42, 0.46	0.16, 0.23	0.35, 0.40	0.40, 0.44
Crackers	0.2293	0.3715	0.3028	0.2798	1	0.36, 0.41	0.24, 0.30	0.25, 0.31	0.26, 0.31	0.23, 0.29	0.34, 0.40	0.39, 0.43
Deterg.	0.2946	0.3673	0.3337	0.3378	0.3873	1	0.26, 0.32	0.18, 0.25	0.25, 0.31	0.30, 0.37	0.46, 0.51	0.48, 0.53
Hot dogs	0.4599	0.3605	0.2486	0.2798	0.2733	0.2927	1	0.23, 0.28	0.26, 0.32	0.30, 0.36	0.31, 0.36	0.29, 0.35
Ice cream	0.2058	0.2682	0.2517	0.1781	0.2819	0.2214	0.2571	1	0.22, 0.27	0.15, 0.23	0.22, 0.26	0.24, 0.30
Non-cola	0.2710	0.2542	0.1759	0.4384	0.2862	0.2783	0.2955	0.2488	1	0.22, 0.29	0.31, 0.35	0.32, 0.36
Sugar	0.3492	0.4026	0.2624	0.1950	0.2595	0.3364	0.3307	0.1937	0.2483	1	0.33, 0.39	0.31, 0.38
Tissue	0.3743	0.4458	0.3804	0.3787	0.3700	0.4830	0.3357	0.2416	0.3260	0.3667	1	0.62, 0.66
Towels	0.2799	0.4104	0.4219	0.4209	0.4077	0.5021	0.3153	0.2730	0.3415	0.3529	0.6397	1

MVP model. A comparison of either Tables 3 or 4 vs. Table 7 indicates that ignoring unobserved heterogeneity across households overstates the estimated inter-category correlations. This finding is in the same spirit as findings in the brand choice literature that ignoring unobserved heterogeneity across households overstates the estimated serial correlation in the error terms in households' random utilities for brands (Allenby & Lenk, 1994; Keane, 1997). We summarize this finding below.

> *Empirical Finding 2:* Ignoring the effects of unobserved heterogeneity across households in the proposed multivariate probit model leads one to overestimate correlations in households' purchase incidence decisions across product categories.

In Tables 8 and 9 we summarize the estimated covariate effects for the twelve product categories based on the six model specifications. While the posterior means are reported in Table 8, the posterior credibility intervals are reported in Table 9. The second column in each table lists the results based on the proposed model estimated on the full set of twelve product categories (i.e. model 1). The estimates of the marketing mix coefficients and product inventory are signed as expected for all twelve categories. Specifically, the coefficients of price are always negative, the coefficients of display and feature are always positive and the coefficients of inventory are always negative. Among the twelve categories, cola beverages show maximum responsiveness to price (posterior mean of –2.1378), ice cream shows maximum responsiveness to store displays (posterior mean of 1.3978), while coffee shows maximum responsiveness to newspaper feature advertising (posterior mean of 1.0471).

The third column of Tables 8 and 9 lists the results based on the proposed model with the unobserved heterogeneity distribution restricted to be common across the twelve product categories (i.e. model 2). A comparison of the estimates in columns 2 and 3 (i.e. model 1 vs. model 2) indicates that household sensitivity to price and display are, by and large,[6] understated in model 2. In other words, restricting the unobserved heterogeneity distribution to be common across product categories leads one to conclude that households are less responsive to pricing and display activities. The feature coefficient, however, shows mixed results, i.e. it is understated for five categories and overstated for the remaining seven categories.

The fourth column of Tables 8 and 9 lists the results based on the proposed model estimated on three mutually exclusive subsets of four product categories (i.e. model 3). A comparison of the estimates in columns 3 and 4 (i.e. model 2 vs. model 3) indicates that household sensitivity to price, display and feature is, by and large,[7] overstated in model 3. Taken together with our earlier findings

Table 8. Posterior Means of Estimates of Covariate Effects.

Parameter	**Model 1** MVP-12 categories Panel (Flexible)	**Model 2** MVP-12 categories Panel	**Model 3** MVP-4 categories Panel	**Model 4** MVP-2 categories[15] Panel	**Model 5** UVP-1 category Panel	**Model 6** MVP-12 categories Cross-sectional
Bacon Intercept	−1.0017	−1.0816	−1.0025	−1.0890	−0.9553	−1.1214
Bacon Price	−1.2178	−0.9810	−1.1417	−1.1282	−1.3549	−0.8619
Bacon Display	0.5153	0.4904	0.5828	0.5062	0.5994	0.4726
Bacon Feature	0.3067	0.3891	0.3440	0.3912	0.2789	0.3746
Bacon Inventory	−0.0398	−0.0438	−0.0446	−0.0372	−0.0381	−0.0412
Butter Intercept	−1.2751	−1.2545	−1.2835	−1.2964	−1.2963	−1.2292
Butter Price	−0.8621	−0.2619	−0.4032	−0.4232	−1.0287	−0.1965
Butter Display	0.4809	0.4697	0.5164	0.5039	0.5141	0.4230
Butter Feature	0.7664	0.8624	0.8779	0.8675	0.7763	0.7858
Butter Inventory	−0.0442	−0.0425	−0.0399	−0.0408	−0.0431	−0.0400
Coffee Intercept	−1.7580	−1.6988	−1.7452	−2.2334	−1.7828	−1.6431
Coffee Price	−0.1302	−0.0056	−0.0029	−1.7977	−0.1683	−0.0039
Coffee Display	0.5877	0.4991	0.5231	0.7222	0.6104	0.4490
Coffee Feature	1.0471	1.0185	1.0697	0.2439	1.0779	0.9091
Coffee Inventory	−0.0268	−0.0308	−0.0290	−0.0037	−0.0266	−0.0285

Table 8. (Continued).

Parameter	Model 1 MVP-12 categories Panel (Flexible)	Model 2 MVP-12 categories Panel	Model 3 MVP-4 categories Panel	Model 4 MVP-2 categories[16] Panel	Model 5 UVP-1 category Panel	Model 6 MVP-12 categories Cross-sectional
Cola Intercept	−2.4828	−1.9734	−2.1031	−2.2340/−2.1801/−1.8149/2.2293	−2.6756	−1.8723
Cola Price	−2.1378	−1.3899	−1.6094	−1.8062/−1.5488/0.0022/−1.8067	−2.4360	−1.2687
Cola Display	0.7405	0.6995	0.7026	0.7228/0.7915/0.5408/0.7253	0.7826	0.6623
Cola Feature	0.2136	0.2341	0.2529	0.2238/0.2615/1.1369/0.2467	0.2081	0.1868
Cola Inventory	−0.0043	−0.0041	−0.0038	−0.0036/−0.0036/−0.0281/0.0037	−0.0041	−0.0039
Crackers Inter.	−1.4098	−1.3874	−1.4256	−1.3257	−1.4504	−1.4210
Crackers Price	−0.0566	−0.0015	0.0140	−0.1612	−0.0560	0.0674
Crackers Display	0.8078	0.7027	0.7464	0.8123	0.8742	0.6522
Crackers Feature	0.5128	0.4576	0.4590	0.5216	0.5618	0.4408
Crackers Invent.	−0.0537	−0.0562	−0.0514	−0.0480	−0.0467	−0.0562
Deterg. Intercept	−1.8150	−1.7058	−1.7350	−1.8472/−1.8614	−1.8600	−1.6292
Deterg. Price	−0.9404	−0.5361	−0.5708	−0.6082/−0.8654	−1.1270	−0.4770
Deterg. Display	0.7598	0.7244	0.7763	0.8224/0.8204	0.8165	0.6462
Deterg. Feature	0.8500	0.8493	0.8740	0.9606/0.9132	0.8444	0.7561
Deterg. Inventory	−0.0118	−0.0113	−0.0103	−0.0106/−0.0106	−0.0100	−0.0110

Table 8. (Continued).

Parameter	Model 1 MVP-12 categories Panel (Flexible)	Model 2 MVP-12 categories Panel	Model 3 MVP-4 categories Panel	Model 4 MVP-2 categories[17] Panel	Model 5 UVP-1 category Panel	Model 6 MVP-12 categories Cross-sectional
Hot dogs Inter.	−0.7690	−1.0369	−1.0198	−0.8788	−0.7476	−1.0735
Hot dogs Price	−1.0381	−0.5991	−0.6643	−0.8951	−1.1282	−0.5138
Hot dogs Display	0.6708	0.7109	0.7254	0.7580	0.7449	0.6808
Hot dogs Feature	0.5043	0.6012	0.6539	0.5394	0.5267	0.5591
Hot dogs Invent.	−0.0212	−0.0218	−0.0230	−0.0210	−0.0221	−0.0209
Ice cream Inter.	−1.8438	−1.5724	−1.6193	NA	−1.8740	−1.5335
Ice cream Price	−0.8738	−0.3546	−0.4206	NA	−0.9170	−0.2945
Ice cream Display	1.3978	1.4967	1.5398	NA	1.4796	1.3689
Ice cream Feature	0.8459	0.9286	0.9439	NA	0.8576	0.8847
Ice cream Invent.	−0.0068	−0.0085	−0.0079	NA	−0.0064	−0.0087
Non-cola Inter.	−1.4032	−1.2917	−1.3256	−1.4185	−1.4230	−1.2653
Non-cola Price	−0.2906	−0.3977	−0.3551	−0.4391	−0.3162	−0.3672
Non-cola Display	0.5337	0.4406	0.5002	0.5240	0.5797	0.4237
Non-cola Feature	0.4000	0.3188	0.3494	0.2939	0.3829	0.2734
Non-cola Invent.	−0.0001	0.0000	−0.0003	0.0000	−0.0002	0.0000

Table 8. (Continued).

Parameter	**Model 1** MVP-12 categories Panel (Flexible)	**Model 2** MVP-12 categories Panel	**Model 3** MVP-4 categories Panel	**Model 4** MVP-2 categories[18] Panel	**Model 5** UVP-1 category Panel	**Model 6** MVP-12 categories Cross-sectional
Sugar Intercept	−2.2659	−2.1551	−2.2408	−2.3497/−2.3294	−2.3595	−2.0453
Sugar Price	−0.2970	−0.3212	−0.3215	−0.4646/−0.3783	−0.4029	−0.2608
Sugar Display	0.6097	0.5896	0.6522	0.6049/0.6234	0.5411	0.5516
Sugar Feature	0.7451	0.6891	0.7074	0.6870/0.7417	0.7984	0.6316
Sugar Invent.	−0.0282	−0.0328	−0.0323	−0.0310/−0.0309	−0.0280	−0.0319
Tissue Inter.	−2.7379	−1.7428	−1.9260	−2.0770/−2.1892	−3.0543	−1.6315
Tissue Price	−0.9703	−0.2505	−0.3523	−0.4274/−0.4962	−1.1890	−0.2043
Tissue Display	0.7121	0.6936	0.7371	0.7819/0.7854	0.7749	0.6303
Tissue Feature	0.8078	0.8388	0.8819	0.9358/0.9241	0.8428	0.7489
Tissue Invent.	−0.0094	−0.0090	−0.0087	−0.0090/−0.0087	−0.0090	−0.0085
Towels Inter.	−2.1006	−1.6328	−1.6993	−1.8862/−1.8505	−2.1975	−1.5480
Towels Price	−1.4211	−0.4173	−0.3848	−0.8690/−0.5492	−1.5782	−0.3431
Towels Display	0.8045	0.7407	0.7933	0.8524/0.8361	0.8771	0.6629
Towels Feature	0.7676	0.8020	0.8736	0.8505/0.9273	0.8069	0.7168
Towels Invent.	−0.0241	−0.0296	−0.0282	−0.0243/−0.0272	−0.0207	−0.0284

Table 9. Posterior Credibility Intervals of Estimates of Covariate Effects.

Parameter	Model 1 MVP-12 categories Panel (Flexible)	Model 2 MVP-12 categories Panel	Model 3 MVP-4 categories Panel	Model 4 MVP-2 categories[19] Panel	Model 5 UVP-1 category Panel	Model 6 MVP-12 categories Cross-sectional
Bacon Intercept	−1.0936, −0.9041	−1.1504, −1.0067	−1.0850, −0.9142	−1.1815, −0.9959	−1.0603, −0.8482	−1.1798, −1.0588
Bacon Price	−1.3062, −1.1244	−1.0551, −0.9105	−1.2169, −1.0725	−1.2156, −1.0493	−1.4506, −1.2702	−0.9351, −0.7947
Bacon Display	0.4311, 0.5975	0.4128, 0.5678	0.5087, 0.6615	0.4076, 0.5887	0.5108, 0.6935	0.4004, 0.5470
Bacon Feature	0.2328, 0.3826	0.3227, 0.4561	0.2734, 0.4174	0.3181, 0.4660	0.2072, 0.3585	0.3100, 0.4327
Bacon Inventory	−0.0488, −0.0312	−0.0535, −0.0353	−0.0543, −0.0349	−0.0456, −0.0281	−0.0474, −0.0282	−0.0495, −0.0325
Butter Intercept	−1.3414, −1.2129	−1.3010, −1.2071	−1.3361, −1.2304	−1.3569, −1.2364	−1.3771, −1.2174	−1.2479, −1.2104
Butter Price	−0.9395, −0.7822	−0.3033, −0.2180	−0.4470, −0.3530	−0.4864, −0.3623	−1.1203, −0.9439	−0.2369, −0.1573
Butter Display	0.4022, 0.5601	0.4007, 0.5402	0.4392, 0.5934	0.4286, 0.5782	0.4280, 0.5990	0.3577, 0.4855
Butter Feature	0.6994, 0.8368	0.8046, 0.9198	0.8193, 0.9400	0.8059, 0.9332	0.7077, 0.8432	0.7297, 0.8421
Butter Inventory	−0.0501, −0.0383	−0.0487, −0.0366	−0.0462, −0.0331	−0.0475, −0.0345	−0.0495, −0.0363	−0.0460, −0.0341
Coffee Intercept	−1.8316, −1.6868	−1.7509, −1.6433	−1.8048, −1.6834	−1.8831, −1.7423	−1.8692, −1.6986	−1.6770, −1.6103
Coffee Price	−0.1903, −0.0654	−0.0423, 0.0287	−0.0409, 0.0360	−0.0394, 0.0433	−0.2320, −0.1062	−0.0385, 0.0305
Coffee Display	0.4999, 0.6714	0.4225, 0.5732	0.4447, 0.6044	0.4560, 0.6278	0.5260, 0.7011	0.3795, 0.5246
Coffee Feature	0.9643, 1.1311	0.9429, 1.0947	0.9814, 1.1509	1.0565, 1.2234	0.9924, 1.1633	0.8374, 0.9789
Coffee Inventory	−0.0314, −0.0222	−0.0352, −0.0256	−0.0345, −0.0241	0.0331, −0.0234	−0.0312, −0.0218	−0.0333, −0.0240

Table 9. (Continued).

Parameter	Model 1 MVP-12 categories Panel (Flexible)	Model 2 MVP-12 categories Panel	Model 3 MVP-4 categories Panel	Model 4 MVP-2 categories[20] Panel	Model 5 UVP-1 category Panel	Model 6 MVP-12 categories Cross-sectional
Cola Intercept	−2.5836, −2.3924	−2.0267, −1.9162	−2.1649, −2.0367	−2.3, −2.2; −2.3, −2.1;2.3, −2.2;2.3, −2.2	−2.7976, −2.5577	−1.9040, −1.8395
Cola Price	−2.2646, −2.0206	−1.4693, −1.3144	−1.7015, −1.5141	−1.9, −1.7; −1.6, −1.4; −1.9, −1.7; −1.9, −1.7	−2.5667, −2.3019	−1.3414, −1.1975
Cola Display	0.6671, 0.8141	0.6380, 0.7634	0.6408, 0.7673	0.6, 0.8; 0.7, 0.9; 0.6, 0.8; 0.7, 0.8	0.7095, 0.8606	0.6065, 0.7160
Cola Feature	0.1453, 0.2770	0.1753, 0.2908	0.1887, 0.3167	0.2, 0.3; 0.2, 0.3; 0.2, 0.3; 0.2, 0.3	0.1346, 0.2794	0.1329, 0.2417
Cola Inventory	−0.0056, −0.0030	−0.0052, −0.0028	−0.0050, −0.0025	−0.0, −0.0; −0.0, −0.0; −0.0, −0.0; −0.0, −0.0	−0.0054, −0.0028	−0.0051, −0.0027
Crackers Inter.	−1.5041, −1.3075	−1.4555, −1.3241	−1.4994, −1.3534	−1.4155, −1.2336	−1.5480, −1.3577	−1.4656, −1.3743
Crackers Price	−0.1249, 0.0074	−0.0497, 0.0452	−0.0403, 0.0671	−0.2202, −0.1039	−0.1240, 0.0124	0.0240, 0.1095
Crackers Display	0.7322, 0.8820	0.6313, 0.7818	0.6762, 0.8219	0.7319, 0.8935	0.7901, 0.9561	0.5877, 0.7177
Crackers Feature	0.4135, 0.6100	0.3728, 0.5411	0.3707, 0.5505	0.4226, 0.6180	0.4641, 0.6604	0.3554, 0.5216
Crackers Invent.	−0.0641, −0.0437	−0.0672, −0.0458	−0.0618, −0.0414	−0.0588, −0.0370	−0.0570, −0.0359	−0.0658, −0.0465
Deterg. Intercept	−1.8841, −1.7476	−1.7540, −1.6553	−1.7859, −1.6813	(−1.92, −1.78) (−1.94, −1.78)	−1.9389, −1.7821	−1.6509, −1.6077
Deterg. Price	−1.0473, −0.8332	−0.6024, −0.4810	−0.6392, −0.5003	(−0.69, −0.53) (−0.95, −0.79)	−1.2451, −1.0198	−0.5380, −0.4157
Deterg. Display	0.6813, 0.8342	0.6503, 0.8098	0.6952, 0.8527	(0.74, 0.90) (0.74, 0.90)	0.7329, 0.9024	0.5777, 0.7114
Deterg. Feature	0.7578, 0.9449	0.7689, 0.9409	0.7916, 0.9571	(0.87, 1.06) (0.82, 1.02)	0.7488, 0.9419	0.6777, 0.8374
Deterg. Inventory	−0.0141, −0.0095	−0.0133, −0.0092	−0.0124, −0.0081	(−0.01, −0.00) (−0.01, −0.00)	−0.0122, −0.0077	−0.0130, −0.0090

Table 9. (Continued).

Parameter	Model 1 MVP-12 categories Panel (Flexible)	Model 2 MVP-12 categories Panel	Model 3 MVP-4 categories Panel	Model 4 MVP-2 categories[21] Panel	Model 5 UVP-1 category Panel	Model 6 MVP-12 categories Cross-sectional
Hot dogs Inter.	−0.8627, −0.6686	−1.1039, −0.9672	−1.0960, −0.9466	−0.9733, −0.7849	−0.8546, −0.6407	−1.1225, −1.0250
Hot dogs Price	−1.1229, −0.9584	−0.6527, −0.5465	−0.7220, −0.6048	−0.9689, −0.8184	−1.2101, −1.0403	−0.5641, −0.4636
Hot dogs Display	0.5719, 0.7660	0.6227, 0.8051	0.6353, 0.8207	0.6574, 0.8602	0.6451, 0.8478	0.5980, 0.7583
Hot dogs Feature	0.43467, 0.5739	0.5448, 0.6649	0.5903, 0.7189	0.4763, 0.6062	0.4553, 0.6031	0.5041, 0.6179
Hot dogs Invent.	−0.0279, −0.0144	−0.0295, −0.0145	−0.0299, −0.0156	−0.0281, −0.0136	−0.0291, −0.0147	−0.0287, −0.0133
Ice cream Inter.	−1.9251, −1.7616	−1.6219, −1.5214	−1.6722, −1.5667	NA	−1.9596, −1.7911	−1.5580, −1.5101
Ice cream Price	−0.9517, −0.7956	−0.3988, −0.3060	−0.4739, −0.3715	NA	−0.9997, −0.8368	−0.3348, −0.2538
Ice cream Display	0.8224, 2.0111	0.9359, 2.0117	0.9918, 2.1033	NA	0.8719, 2.0769	0.8542, 1.8706
Ice cream Feature	0.7816, 0.9059	0.8730, 0.9851	0.8890, 1.0004	NA	0.7946, 0.9239	0.8280, 0.9378
Ice cream Invent.	−0.0087, −0.0048	−0.0104, −0.0067	−0.0098, −0.0058	NA	−0.0083, −0.0045	−0.0106, −0.0068
Non-cola Inter.	−1.4893, −1.3179	−1.3422, −1.2400	−1.3842, −1.2691	−1.4994, −1.3454	−1.5053, −1.3408	−1.2904, −1.2397
Non-cola Price	−0.3571, −0.2168	−0.4507, −0.3453	−0.4132, −0.2961	−0.5038, −0.3761	−0.3900, −0.2451	−0.4156, −0.3202
Non-cola Display	0.4617, 0.6038	0.3789, 0.4996	0.4344, 0.5649	0.4561, 0.5909	0.5076, 0.6502	0.3677, 0.4832
Non-cola Feature	0.3235, 0.4776	0.2467, 0.4020	0.2683, 0.4302	0.2180, 0.3714	0.3038, 0.4623	0.2047, 0.3441
Non-cola Invent.	−0.0009, 0.0008	−0.0009, 0.0009	−0.0012, 0.0007	−0.0008, 0.0009	−0.0010, 0.0007	−0.0009, 0.0008

Table 9. (Continued).

Parameter	Model 1 MVP-12 categories Panel (Flexible)	Model 2 MVP-12 categories Panel	Model 3 MVP-4 categories Panel	Model 4 MVP-2 categories[22] Panel	Model 5 UVP-1 category Panel	Model 6 MVP-12 categories Cross-sectional
Sugar Intercept	−2.3604, −2.1714	−2.2283, −2.0822	−2.3219, −2.1594	−2.44, −2.25; −2.42, −2.23	−2.4612, −2.2670	−2.1016, −1.9880
Sugar Price	−0.3816, −0.2132	−0.3950, −0.2515	−0.3944, −0.2480	−0.54, −0.38; −0.46, −0.29	−0.4884, −0.3158	−0.3241, −0.1961
Sugar Display	0.5162, 0.6986	0.5097, 0.6664	0.5723, 0.7352	0.51, 0.69; −0.53, 0.71	0.4659, 0.6163	0.4813, 0.6273
Sugar Feature	0.6569, 0.8359	0.6028, 0.7703	0.6206, 0.7916	0.59, 0.78; 0.66, 0.84	0.7155, 0.8860	0.5564, 0.7064
Sugar Invent.	−0.0328, −0.0238	−0.0373, −0.0283	−0.0372, −0.0273	−0.03, −0.02; −0.04, −0.03	−0.0324, −0.0235	−0.0368, −0.0272
Tissue Inter.	−2.8706, −2.5903	−1.8083, −1.6740	−2.0100, −1.8381	−2.18, −1.97; −2.29, −2.09	−3.2222, −2.8977	−1.6802, −1.5830
Tissue Price	−1.0593, −0.8718	−0.2998, −0.2002	−0.4086, −0.2978	−0.49, −0.36; −0.56, −0.44	−1.3030, −1.0819	−0.2468, −0.1616
Tissue Display	0.6394, 0.7813	0.6282, 0.7601	0.6712, 0.8020	0.71, 0.85; 0.71, 0.86	0.7031, 0.8511	0.5711, 0.6872
Tissue Feature	0.7383, 0.8762	0.7714, 0.8985	0.8165, 0.9501	0.86, 1.01; 0.86, 0.99	0.7677, 0.9152	0.6896, 0.8091
Tissue Invent.	−0.0112, −0.0075	−0.0107, −0.0071	−0.0106, −0.0068	−0.01, −0.00; −0.01, −0.00	−0.0109, −0.0072	−0.0100, −0.0069
Towels Inter.	−2.2015, −1.9771	−1.6806, −1.5804	−1.7575, −1.6413	−1.97, −1.81; −1.94, −1.77	−2.3115, −2.0882	−1.5722, −1.5223
Towels Price	−1.5576, −1.2715	−0.4939, −0.3420	−0.4601, −0.3055	−0.97, −0.78; −0.63, −0.46	−1.7236, −1.4464	−0.4050, −0.2757
Towels Display	0.7374, 0.8719	0.6797, 0.8045	0.7285, 0.8593	0.78, 0.93; 0.77, 0.91	0.8022, 0.9502	0.6071, 0.7193
Towels Feature	0.6947, 0.8425	0.7397, 0.8644	0.8048, 0.9435	0.77, 0.93; 0.85, 1.00	0.7287, 0.8851	0.6607, 0.7729
Towels Invent.	−0.0298, −0.0183	−0.0358, −0.0238	−0.0340, −0.0222	−0.03, −0.02; −0.03, −0.02	−0.0266, −0.0148	−0.0337, −0.0233

that model 3 understates correlations across product categories, this means that the marketing mix variables bear the burden of explaining purchase incidence decisions in product categories that are, in part, due to inter-category correlations (that are incompletely accounted for in the model).

The fifth column in Tables 8 and 9 lists the results from a bivariate version of the proposed model estimated separately on seven different pairs of product categories (i.e. model 4). A comparison of the estimates in columns 3 and 5 (i.e. model 2 vs. model 4) indicates that household sensitivity to price, display and feature is, by and large,[8] overstated in model 4. This finding is consistent with that obtained from comparing models 2 and 3, as discussed in the previous paragraph.

The sixth column in Tables 8 and 9 lists the results from univariate binary probit models estimated separately for the twelve product categories (i.e. model 5). A comparison of the estimates in columns 3 and 6 (i.e. model 2 vs. model 5) indicates that household sensitivity to price, display and feature is, by and large,[9] overstated in model 5. This finding, consistent with the findings obtained by comparing either models 2 and 3 or models 2 and 4, is summarized below.

> *Empirical Finding 3:* A limited operationalization of the proposed multivariate probit model, using a subset of the full set of product categories within the household's shopping basket (as in Chintagunta & Haldar, 1998; Manchanda, Ansari & Gupta, 1999), leads one to overestimate the effects of marketing variables on households' purchase incidence decisions within each product category.

The seventh column in Tables 8 and 9 lists the results from a purely cross-sectional version – one that ignores unobserved heterogeneity across households – of the proposed multivariate probit model (i.e. model 6). A comparison of the estimates in columns 3 and 7 (i.e. model 2 vs. model 6) seems to indicate that household sensitivity to price, display and feature are overstated in model 6. However, such an interpretation must be kept in check on account of a scale incompatibility problem while comparing models 2 and 6, since the cross-sectional probit (i.e. model 6) does not accommodate random effects across households.

7. SUMMARY

We propose a multivariate probit model with unobserved heterogeneity to explain households' purchase incidence decisions simultaneously across all product categories within their shopping baskets. We estimate the proposed model using basket-level purchase data on a scanner panel of 300 households. We find that a limited operationalization of the proposed model, using a subset of the full set of product categories within the household's shopping basket, leads one to underestimate inter-category correlations and overestimate the effectiveness of marketing variables. We also find that ignoring unobserved heterogeneity across households leads one to overestimate inter-category correlations and underestimate the effectiveness of marketing variables.

One obvious managerial benefit of our proposed model is that retailer can design optimal prices simultaneously across all product categories, taking cross-category correlations into account, in order to maximize store profits. When cross-category correlations exist, ignoring their effects and maximizing category profits independently across product categories will lead to sub-optimal profits. While the findings of this paper are of managerial interest in and of themselves, the implications of these findings on related household decisions, such as brand choice, are of managerial interest. We are currently extending our proposed model to accommodate households' brand choice decisions within each product category. In this framework, we employ a multinomial logit model for households' conditional brand choices within each product category, coupled with a multivariate probit model of households' purchase incidence decisions across product categories. Whether our reported findings about cross-category correlations in purchase incidence decisions in this paper generalize to such a fully specified framework is an area of ongoing investigation.

Last, but not the least, it will be useful to accommodate unobserved heterogeneity along multiple dimensions (instead of in the intercept term only) and model correlations not only in the error terms but also in household response parameters across product categories. This will allow us to investigate whether households exhibit similar sensitivities to the marketing variables in different product categories using a basket-level analysis (Seetharaman, Ainslie & Chintagunta, 1999 investigate this issue using conditional brand choice data on a panel of households in five product categories).

NOTES

1. To the extent that product categories within a household's shopping basket vie for a limited shopping budget of the household, the budget constraint induces cross-category dependencies as well.
2. In our application, $J=12$, $k=5$ which makes the total number of estimated parameters 138.
3. The excluded product categories are barbecue sauce, cat food, cereals, cleansers, cookies, eggs, nuts, pills, pizza, snacks, soap, softener, yogurt.
4. In a companion paper, in which we model both category purchase incidence and brand choice, we explicitly investigate the consequences of such aggregation on model-based inferences.

5. We thank an anonymous reviewer for alerting us to this issue.

6. Except for display coefficients for hot dogs and ice cream, and price coefficients for non-cola and sugar, this holds for the remaining twenty coefficients.

7. Except for bacon's feature coefficient and the price coefficients for coffee, non-cola and towels, this holds for the remaining thirty-two marketing mix coefficients.

8. This holds for forty-eight out of the fifty-four marketing mix variables in tables 8–11.

9. The overstatement holds for twenty-eight out of thirty-six coefficients.

10. Lower and upper halves of the matrix contain posterior means and credibility intervals respectively.

11. Lower and upper halves of the matrix contain posterior means and credibility intervals respectively.

12. Lower and upper halves of the matrix contain posterior means and credibility intervals respectively.

13. Credibility Interval

14. Lower and upper halves of the matrix contain posterior means and credibility intervals respectively.

15. The estimates of bacon and butter are based on hot dogs and sugar as the respective second categories.

16. The four sets of estimates for cola are based on sugar, non-cola, coffee and crackers respectively as the second category. The estimates for crackers are based on cola as the second category. The two sets of estimates for detergents are based on tissue and towels respectively as the second category.

17. The estimates of hot dogs and non-cola are based on bacon and cola as the respective second categories.

18. The two sets of estimates of sugar are based on butter and cola respectively as the second category. The two sets of estimates for tissue are based on detergents and towels respectively as the second category. The two sets of estimates for towels are based on detergents and tissue respectively as the second category.

19. The estimates of bacon and butter are based on hot dogs and sugar as the respective second categories.

20. The four sets of estimates for cola are based on sugar, non-cola, coffee and crackers respectively as the second category. The estimates for crackers are based on cola as the second category. The two sets of estimates for detergents are based on tissue and towels respectively as the second category.

21. The estimates of hot dogs and non-cola are based on bacon and cola as the respective second categories.

22. The two sets estimates of sugar are based on butter and cola respectively as the second category. The two sets of estimates for tissue are based on detergents and towels respectively as the second category. The two sets of estimates for towels are based on detergents and tissue respectively as the second category.

ACKNOWLEDGMENTS

We thank marketing seminar participants at the John M. Olin School of Business at Washington University for their comments on an earlier version of this paper.

REFERENCES

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

Allenby, G. M., & Lenk, P. J. (1994). Modeling Household Purchase Behavior with Logistic Normal Regression. *Journal of the American Statistical Association, 89*(428), 1–14.

Ashford, J. R., & Sowden, R. R. (1970). Multivariate Probit Analysis. *Biometrics, 26*, 535–546.

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

Berry, J. A., & Linoff, G. (1997). *Data Mining Techniques*. Wiley and Sons.

Bucklin, R. E., & Lattin, J. (1991). A Two-State Model of Purchase Incidence and Brand Choice. *Marketing Science, 10*(1), 24–39.

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

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

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

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

Gelfand, A. E., & Smith, A. F. M. (1990). Sampling-Based Approaches to Calculating Marginal Densities. *Journal of the American Statistical Association, 85*(410), 398–409.

Gonul, F., & Srinivasan, K. (1993). Modeling Multiple Sources of Heterogeneity in Multinomial Logit Models: Methodological and Managerial Issues. *Marketing Science, 12*(3), 213–229.

Keane, M. P. (1997). Modeling Heterogeneity and State Dependence in Consumer Choice Behavior. *Journal of Business and Economic Statistics, 15*(3), 310–327.

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

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

Tanner, M. A., & Wong, W. H. (1987). The Calculation of Posterior Distributions by Data Augmentation. *Journal of the American Statistical Association, 82*(398), 528–540.

Tierney, L. (1987). Markov Chains for Exploring Posterior Distributions. *Annals of Statistics, 22*, 1701–1762.