Quantitative Marketing and Economics, 1, 425–465, 2003.

Leveraging Information Across Categories

RAGHURAM IYENGAR*
Graduate School of Business, Uris Hall, Columbia University, 3022 Broadway, New York, NY 10027
E-mail: ri75@columbia.edu

ASIM ANSARI
Graduate School of Business, Uris Hall, Columbia University, 3022 Broadway, New York, NY 10027
E-mail: maa48@columbia.edu

SUNIL GUPTA
Graduate School of Business, Uris Hall, Columbia University, 3022 Broadway, New York, NY 10027
E-mail: sg37@columbia.edu

Abstract. Companies are collecting increasing amounts of information about their customers. This effort is based on the assumption that more information is better and that this information can be leveraged to predict customers' behavior in a variety of situations and product categories. For example, information about a customer's purchase behavior in one category can be helpful in predicting his potential behavior in a related category, which in turn could help a firm in its cross-selling efforts.

In this paper, we present a model to better understand and predict a consumer's purchases and preferences when we may have limited or no information about him in one or more product categories. Conceptually this involves leveraging information from purchases of other consumers in multiple categories as well as partial information (e.g., purchase in one of the categories) of the target consumer. Our approach builds on the pioneering work of Rossi et al. (1996) who demonstrate the value of purchase information in the context of a single product category. We present results from an extensive simulation as well as an application on scanner panel data.

Our simulation shows many interesting and somewhat surprising results. Specifically, we find that compared to a single-category analysis, a cross-category analysis does not lead to any significant improvement in data likelihood in most cases. Therefore, the single-category analysis of Rossi et al. (1996) is even more powerful than previously thought. However, we also find that a cross-category analysis does improve parameter recovery in many situations as compared to a single-category analysis. It is in these conditions that retailers can use cross-category information to better implement micro marketing programs.

We demonstrate the transfer of information across categories in an application of two grocery products—Breakfast Foods and Table Syrup. In spite of a reasonable correlation (0.21) in the price parameter across these two categories, our simulation guidelines predict very little benefit of cross-category analysis over single-category analysis. Our empirical results confirm this prediction.

Key words. cross category analysis, information transfer, hierarchical bayes

JEL Classification: C11, C15, C23, C33, C35

*Corresponding author.

1. Introduction

Understanding and predicting consumer purchases is one of the main goals of marketing researchers and marketing managers. Several decades of research has shown that the best predictor of a consumer's purchase is his past purchase behavior. The growing realization that past purchases are important predictors, the ease of obtaining purchasing data and the advent of sophisticated information technology has propelled many companies to create enormous databases for customer relationship management (CRM) and customer targeting.[1]

Even with these large databases, it is not uncommon, however, to have limited purchase information about a specific consumer. This may be because of long purchase cycles for a product, or because of missing information arising out of a lack of data integration across various information sources (e.g., call centers, web sites, retail stores, etc.) or owing to less than comprehensive information for newly acquired customers. Such data scarcity poses challenges in defining a customer-specific profile (e.g., individual customer's price sensitivity) for targeting purposes. Recently researchers, especially Rossi et al. (1996), have made significant progress in addressing customer targeting when partial data is available on some customers. This research shows how Bayesian methods can be used to optimally pool information across consumer purchases within a product category. Such Bayesian information pooling enables one to estimate an individual customer's preference (for example, price sensitivity) with very limited or no information.

In a related research stream, several researchers have examined the correlation in consumers' price sensitivities across multiple product categories (Ainslie and Rossi, 1998; Erdem et al., 2001). This research shows that consumer price sensitivities are correlated across product categories and that a consumer's price sensitivity in one category provides strong indications about his price sensitivity in another category. Researchers typically arrive at this conclusion by using a combination of customer demographics and information about previous brand choices and their causal context (i.e., full information) from multiple product categories.

There are many situations where a company does not have complete information about consumers' purchases across multiple categories, thus making it difficult to directly apply the methods developed in previous research (Ainslie and Rossi, 1998, and others). For example, in cross-selling situations (Kasulis et al., 1979; Kamakura et al., 2003) a firm may have significant purchase information about a customer in one product category but may possess no information about him in a related category. Clearly methods are needed that can handle such incomplete data situations across categories. As customer acquisition becomes more and more expensive, a firm's ability to better target its customers for cross-selling purposes gains more importance. Therefore, there is a strong need to better leverage

1 We are using the terms consumers, customers and households interchangeably.

consumers' purchase information in one category to make inferences about their potential purchase behavior in other product categories.

In this paper, we address such situations involving partial information availability across product categories. Specifically, we propose methods to better understand and predict a consumer's purchases when we may have limited or no information about him in one or more categories. Conceptually this involves leveraging information from other consumers across the multiple categories as well as information about the purchase behavior of our target customer in other categories. Therefore, our approach builds on the work of Rossi et al. (1996) who demonstrate the value of information within a product category, and Ainslie and Rossi (1998) and Erdem (1998) who show how complete information about consumer purchases across multiple categories can be used to estimate cross-category linkages. Unlike Rossi et al. (1996) we propose a method to leverage information across multiple categories, and unlike Ainslie and Rossi (1998) we deal with situations, such as cross-selling, where we have incomplete information about consumer purchases in one or more categories. Our work is also conceptually related to research on data fusion (Wedel and Kamakura, 1997) that focuses on combining information from different data sources.

Does it always help a firm to utilize information from multiple categories when trying to target customers? Intuition would suggest that more information is always better. However, recent industry reports on CRM suggest that collecting data from multiple sources and especially integrating these data is both complex and costly. If the marginal benefit of additional information is limited, then this additional cost and complexity may not be warranted. To address this issue we use a simulation and show that there are limits to the value of cross-category information. Specifically, we show that even when the cross-category correlation between the individual-specific parameters is high, the value of cross-category information may be insignificant under some reasonable situations. In other words, in certain situations, information from a single category is as good as multiple category information in predicting consumer purchase behavior. This result has several implications. First, it makes the approach suggested by Rossi et al. (1996) even more important. Second, it suggests that rather than collecting more data, firms are well advised to identify conditions under which multiple category information is indeed going to be valuable in predicting consumer behavior. In our empirical application we show how we can use the guidelines that emerge from our simulation to anticipate the value of cross-category information.

The rest of the paper is organized as follows: In Section 2, we present our conceptual framework for leveraging information across categories under different Scenarios or levels of available information. In Section 3, we describe our modeling approach and show how Bayesian methods can be used to make inferences about specific customers given partial information across categories. In Section 4, we report the results of a simulation study that identifies the conditions under which information can be transferred across categories. In Section 5, we apply our methods on a two-category data set. In Section 6, we show an extension of our model to multiple categories by considering a three category example. In Section 7, we comment on how our findings can be reconciled with previous research on multi-category data. Finally in Section 8, we conclude.

2. Conceptual framework

We illustrate our conceptual framework with two product categories. An extension to multiple categories is illustrated in Section 6. We also distinguish between two groups of customers—the reference group and the target group. The reference group consists of those customers for whom a firm has complete information on all product categories. Complete information for this group may be obtained in many ways, such as by augmenting company's databases with surveys. For example, a bank's database does not provide any information about its customers' activities with competing banks. Therefore, it lacks information about its customers' share of wallet which is crucial to assess their future potential value for the firm. Kamakura et al. (2003) show how a survey with a sample of customers can be used to augment a bank's database to provide a complete picture for these reference customers. All remaining customers belong to the target group and the firm has incomplete information about them in one or more categories. The objective of the firm is to assess and predict purchase behavior (e.g., price sensitivity) of the target customers in all product categories. These estimates can then be used for targeting (e.g., who to cross-sell) as well as customization (e.g., what price to offer) purposes.

Figure 1 shows the intuition for leveraging information across categories. In the case of a single category analysis, complete information on reference customers in Category-1 is used to estimate their response parameters (e.g., price sensitivity) and to infer the parameters of target households in Category-1. A similar analysis can be done independently for the second category. However, this approach ignores the linkages across categories. For example, consumers' price sensitivities may be highly correlated across product categories. A cross-category analysis attempts to leverage this correlation.

In a cross-category analysis we have complete information on reference consumers about their purchases and causal variables in both categories. However, there is limited information about the target customers. As discussed shortly, there may be various levels of incomplete information for the target customers. For example, in a cross-selling situation, we may have complete information about a target customer in Category-1 but no information about his behavior in Category-2. In order to impute parameter estimates for the target customer in Category-2, we leverage information from multiple sources—information from reference customers in Category-2 (akin to single category analysis), information about the correlation or similarity of purchase behavior across the two categories from the reference group (akin to Ainslie and Rossi, 1998), as well as the information, if any, about the purchase of target customer in Category-1. In the model section, we formalize this intuition under different Scenarios pertaining to different levels of partial information availability for the target customers.

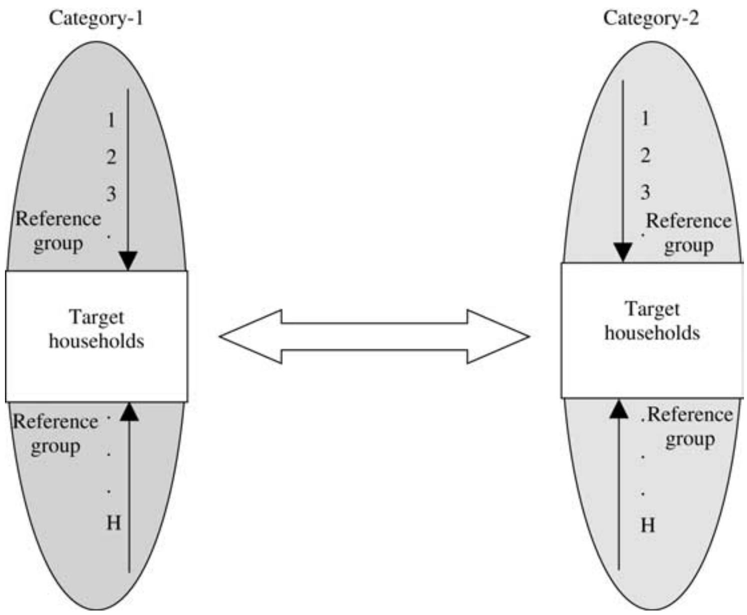

Figure 1. Information transfer between categories.

As the details and the complexity of modeling vary with the degree of information available for target customers, we begin by discussing the different levels of information that firms may have about these customers. Note that throughout this analysis we assume that the firm has complete information for a reference group of customers.

2.1. Information sets

Following Rossi et al. (1996), we assume that within a single product category a firm may have four levels of information about target customers:[2]

1. Base information: In this situation, firms possess no specific information about customers within a product category. They may, however, possess aggregate information about the distribution of demographics in the market place.
2. Demographics: In this situation, firms possess no information about choices made by customers, but have information about the demographics of individual customers. This is very common in direct marketing where firms buy lists of customers with a known demographic profile.

2 Details regarding the situations in which such information sets may arise and the rationale for considering them are discussed in Rossi et al. (1996).

Table 1. Information matrix.

	Category-2			
Category-1	Base	Demo	Choice	Full
Base	Scenario 1	—	—	—
Demo	—	Scenario 2	Scenario 3	Scenario 4
Choice	—	Scenario 3	Scenario 5	Scenario 6
Full	—	Scenario 4	Scenario 6	Scenario 7

3. Choice only: Here, firms have information about the previous brand choices of customers as well as their demographics, but have no information about the causal environment. Many CRM and direct marketing applications can be characterized by this situation.
4. Full purchase history: Full purchase history involves knowledge about the choices, demographics and the causal variables for all purchase observations for all customers. This represents the case with complete information.[3]

Rossi et al. (1996) discussed these information sets for one product category. When considering multiple categories, information sets for each category can be combined to form cross-category information sets for a target customer. For example, for two categories, the combined information sets or Scenarios can be represented in an information matrix as shown in Table 1. It is evident that some Scenarios are infeasible and these are represented with a "—" in Table 1. For example, once we know the demographic information for a target customer, it is available for both categories and therefore, the Scenario involving Base information in Category-1 and Demo information in Category-2 is infeasible.

In Scenario-1 we have only aggregate level information, which makes it difficult to customize offers or target specific customers. On the other extreme is Scenario-7 where we have complete information about all customers in both product categories. Ainslie and Rossi (1998) examine this Scenario to estimate the degree of correlation in parameters across categories. The intermediate Scenarios 2–6 represent varying degrees of information for each of the two product categories. They also represent typical situations in CRM and database marketing. For example, Scenario-4 is a cross-selling situation where we have complete information about a target customer in one category but no information (except his demographics) in the other category. Scenario-3 is an incomplete version of Scenario-4 where the firm did not collect causal information (e.g., what catalogs or offers were sent to target customers). Note, while Rossi et al. (1996) deal with only one product category and Ainslie and

3 Rossi et al. (1996) also consider the case of one observation, which is very similar to the case of full information (or multiple observations), both in terms of concept and modeling. Therefore, we omit this case.

Rossi (1998) and Erdem (1998) deal with only Scenario-7, we develop a model for all information Scenarios and thereby assess the value of additional information for prediction and targeting in this cross-category context. We now develop a formal model and show how Bayesian methods can be used to leverage information across categories for different Scenarios.

3. Model

We use a brand choice context, as in Rossi et al. (1996), to illustrate how information can be leveraged across different product categories. We begin by describing a hierarchical Bayesian cross-category choice model for the reference households for whom we have full information across the different categories. Next, we show how various posterior predictive distributions from this full model can be used to estimate household-specific coefficients for the target households under different information Scenarios.

3.1. Model for reference households

Consider $h = 1, \ldots, H$ households of the reference group who provide $t = 1, \ldots, n_{hk}$ observations for each of $k = 1, \ldots, K$ categories. Each observation in category k yields $\mathbf{y}_{htk}$, a vector of binary indicators $\{y_{h1tk}, \ldots, y_{hjtk}, \ldots, y_{hp_ktk}\}$ that identify the brand chosen in that category at time t. A random utility framework can be used to model the cross-category choices in terms of brand utilities as follows:

$$\mathbf{u}_{htk} = \mathbf{X}_{htk}\boldsymbol{\beta}_{hk} + \mathbf{e}_{htk}, \tag{1}$$

where $\mathbf{X}_{htk}$ denotes the $p_k \times q_k$ matrix containing causal variables in category k, p_k represents the number of utility equations and q_k represents the number of coefficients (intercepts and response coefficients) to be estimated. We assume that the errors within a category $\mathbf{e}_{htk}$ are independent and are distributed $N(0, \boldsymbol{\Sigma}_k)$ where $\boldsymbol{\Sigma}_k$ represents the block diagonal covariance matrix for the unobserved factors. This yields a multinomial probit specification within each category. We also assume that the errors are independent across the categories. The link in the utilities across the categories is achieved by assuming that the household-specific coefficients, $\boldsymbol{\beta}_h = \{\boldsymbol{\beta}_{h1}, \boldsymbol{\beta}_{h2}, \ldots, \boldsymbol{\beta}_{hK}\}$, are jointly distributed multivariate normal. Specifically,

$$\boldsymbol{\beta}_h = \mathbf{Z}_h\boldsymbol{\alpha} + \boldsymbol{\lambda}_h, \tag{2}$$

where $\mathbf{Z}_h$ is a matrix containing the demographic variables for household h, $\boldsymbol{\alpha}$ represents the population mean, and $\boldsymbol{\lambda}_h$, which contains the unobserved sources of consumer heterogeneity, is assumed to be distributed multivariate normal, $N(0, \boldsymbol{\Lambda})$.

Note that the $\mathbf{\Lambda}$ matrix captures the covariance among the parameters across product categories. Specifically,

$$\mathbf{\Lambda} = \begin{bmatrix} \mathbf{\Lambda}_{11} & \mathbf{\Lambda}_{12} & \mathbf{\Lambda}_{13} & \dots & \mathbf{\Lambda}_{1K} \\ & \mathbf{\Lambda}_{22} & \mathbf{\Lambda}_{23} & \dots & \mathbf{\Lambda}_{2K} \\ & & \dots & & \\ & & & & \mathbf{\Lambda}_{KK} \end{bmatrix}, \tag{3}$$

where $\mathbf{\Lambda}_{kk'}$ is the covariance matrix between the parameters of product category k and k'. An independent single category analysis implicitly assumes that each $\mathbf{\Lambda}_{kk'} = \mathbf{0}$ where $k \neq k'$. The link between the latent utilities and the observed outcome is specified in the following manner:

$$y_{hjtk} = \begin{cases} 1 & \text{if } \max\,(\mathbf{u}_{htk}) = u_{hjtk} \\ 0 & \text{otherwise,} \end{cases} \tag{4}$$

where, $j = 1, \dots, p_k$ and $k = 1, \dots, K$.

As the origin and scale of the utilities are unknown, identification is achieved by constraining one intercept to zero and by setting one of the variances in $\mathbf{\Sigma}_k$ to one for each product category. Bayesian analysis can be performed by specifying priors over all unknowns and by using MCMC methods for simulating from the posterior distribution. Details regarding the priors and the full conditional distributions for the unknowns are provided in Appendix A. It can be seen in Appendix A that a cross-category analysis involving two categories requires posterior conditional distributions of $\boldsymbol{\beta}_{h1}$ given $\boldsymbol{\beta}_{h2}$ and $\boldsymbol{\beta}_{h2}$ given $\boldsymbol{\beta}_{h1}$. Here, $\boldsymbol{\beta}_{h1}$ and $\boldsymbol{\beta}_{h2}$ refer to the household-specific parameters for Category-1 and Category-2. These posterior computations explicitly connect the household-specific parameters of the categories and reflect the effect of any cross-category covariance between the parameters, an important aspect that is ignored by a single category analysis. Also, note that there is no restriction in the length of parameter vectors of the two categories—the parameter vector in Category-1 can be of a different length than that in Category-2. This difference can be due to either unequal number of brands in the two categories or different marketing mix elements across the two categories. For instance, consider three brands in Category-1 and four brands in Category-2 with price as the only marketing mix element in both categories. Then, the length of the parameter vector in Category-1 would be three (two brand intercepts and one price parameter). Similarly, the length of the parameter vector in Category-2 would be four (three brand intercepts and one price parameter). Given these parameter vectors, $\mathbf{\Lambda}_{11}$ would be of size 3×3 and $\mathbf{\Lambda}_{22}$ would be a 4×4 matrix. The cross-category covariance, $\mathbf{\Lambda}_{12}$, however, would be of size 3×4. All elements of the cross-category covariance matrix can be estimated. Thus, there is no restriction that the parameter vectors in the two categories have to be of equal length.

Once the above model is estimated on a reference set of individuals, various posterior predictive distributions can be computed to obtain household-specific coefficients for the target households about whom we have only partial information.

3.2. *Model for target households*

We now show how inferences can be made for the household-specific coefficients $\boldsymbol{\beta}_h$ for the target group, under different information Scenarios. For each Scenario, inferences about $\boldsymbol{\beta}_h$ are obtained by combining information from a full information set analysis on a reference group of households with other sources of information that are specific to the target household *h*. In the main body of the paper, we describe how such inferences can be made for two Scenarios. These two Scenarios represent cases which are either typical (e.g., cross selling as in Scenario-4) or involve sparse information (e.g., Scenario-5 represents a situation where a firm lacks information about causal variables for the target households). Complete mathematical details regarding the computation of $\boldsymbol{\beta}_h$ under the other Scenarios are available in Appendix B.

3.2.1. Scenario-4: Full-Demo. This Scenario represents typical cross-selling situations where a firm has complete information about target households in one category (without loss of generalization, say, Category-1) but has no information, except for demographics in another category (Category-2). Inferences about the response coefficients $\boldsymbol{\beta}_{h2}$ in Category-2 can be obtained by combining information from a reference set of households on whom full information in both categories is available, with information about the target household *h* in Category-1. Specifically, we have

$$\begin{aligned} \boldsymbol{\beta}_{h1} &= \mathbf{Z}_{h1}\boldsymbol{\alpha}_1 + \boldsymbol{\lambda}_{h1}, \\ \boldsymbol{\beta}_{h2} &= \mathbf{Z}_{h2}\boldsymbol{\alpha}_2 + \boldsymbol{\lambda}_{h2}, \end{aligned} \tag{5}$$

where $\boldsymbol{\lambda}_h = \{\boldsymbol{\lambda}_{h1}, \boldsymbol{\lambda}_{h2}\} \sim N(0, \boldsymbol{\Lambda})$. The reference households provide information about the population parameters $\boldsymbol{\alpha}$ and $\boldsymbol{\Lambda}$. Note that

$$\boldsymbol{\Lambda} = \begin{bmatrix} \boldsymbol{\Lambda}_{11} & \boldsymbol{\Lambda}_{12} \\ \boldsymbol{\Lambda}_{21} & \boldsymbol{\Lambda}_{22} \end{bmatrix}, \tag{6}$$

where $\boldsymbol{\Lambda}_{11}$ and $\boldsymbol{\Lambda}_{22}$ capture the covariances among the parameters of Category-1 and Category-2 respectively, and $\boldsymbol{\Lambda}_{12} = \boldsymbol{\Lambda}_{21}$ capture the cross-category covariances among the parameters of the two categories. If both categories are modeled independently or if we assume no correlation across the parameters of the two categories, then $\mathbf{\Lambda}_{12} = \mathbf{0}$. In this case, parameters of one category provide no information about the parameters in the other category. In other words, there is no information leverage from one category to the other.

If $\mathbf{\Lambda}_{12} \neq \mathbf{0}$ then we can derive the conditional distribution of $\boldsymbol{\beta}_{h2}$ given $\boldsymbol{\beta}_{h1}$, which can then be used to make inferences about $\boldsymbol{\beta}_{h2}$. Specifically,

$$\begin{aligned} \boldsymbol{\beta}_{h2} \mid \boldsymbol{\beta}_{h1} &\sim N(E(\boldsymbol{\beta}_{h2} \mid \boldsymbol{\beta}_{h1}), \mathrm{Var}(\boldsymbol{\beta}_{h2} \mid \boldsymbol{\beta}_{h1})), \\ E(\boldsymbol{\beta}_{h2} \mid \boldsymbol{\beta}_{h1}) &= \mathbf{Z}_{h2}\boldsymbol{\alpha}_2 + \mathbf{\Lambda}_{21}\mathbf{\Lambda}_{11}^{-1}(\boldsymbol{\beta}_{h1} - \mathbf{Z}_{h1}\boldsymbol{\alpha}_1) \\ \mathrm{Var}(\boldsymbol{\beta}_{h2} \mid \boldsymbol{\beta}_{h1}) &= \mathbf{\Lambda}_{22} - \mathbf{\Lambda}_{21}\mathbf{\Lambda}_{11}^{-1}\mathbf{\Lambda}_{12}. \end{aligned} \tag{7}$$

Note that if there is no correlation in parameters across categories, then the mean of this conditional distribution reduces to $\mathbf{Z}_{h2}\boldsymbol{\alpha}_2$, and its variance becomes $\mathbf{\Lambda}_{22}$. Inference about $\boldsymbol{\beta}_{h2}$ can be based on random draws obtained from the conditional distribution in equation (7). In obtaining the $\boldsymbol{\beta}_{h2}$ draws, the MCMC draws for $\boldsymbol{\alpha}$ and $\mathbf{\Lambda}$ obtained from the reference set analysis are used in combination with the $\boldsymbol{\beta}_{h1}$ draws obtained from the single category analysis on the target group of households. This procedure ensures that inferences for $\boldsymbol{\beta}_{h2}$ are based on its posterior predictive distribution.

3.2.2. Scenario-5: Choice-Choice. Increasingly firms are creating loyalty programs to reward and retain their valuable customers. These programs typically keep detailed information about customers' purchases (e.g., airline tickets, hotel stays, or product purchases). This information is needed to reward loyal customers and encourage their patronage with the firm. However, in many of these situations, firms do not keep track of the causal variables that may have affected a consumer's choice. For example, while an airline may keep track of a customer's ticket purchase for its frequent flier program, it may not record the price he paid for that ticket. Similarly, a catalog company may have an extensive customer database that keeps account of all customers' purchases for RFM analysis, but it may not track information about e-mails, catalogs or other promotional items sent to each customer at each point in time. Scenario-5 represents these situations where firms have information only about consumers' choices in both product categories.

As Rossi et al. (1996) show in the context of a single category purchase, lack of information about causal variables poses a significant challenge. Essentially, we have to estimate a consumer's price sensitivity without knowing the prices faced by this consumer. This complexity is further exacerbated in the context of two or more product categories because we also need to capture the correlation in response sensitivities across product categories. We achieve this by extending the procedure developed by Rossi et al. (1996) to two categories. Specifically, utilities for a complete data model (given in equation (1)) can be rewritten as

$$\mathbf{u}_{ht1} = \mathbf{X}_{ht1}\boldsymbol{\beta}_{h1} + \mathbf{e}_{ht1} = \begin{bmatrix} 0 \\ \boldsymbol{\gamma}_{h1} \end{bmatrix} + \mathbf{X}_{cht1}\boldsymbol{\delta}_{h1} + \mathbf{e}_{ht1},$$
$$\mathbf{u}_{ht2} = \mathbf{X}_{ht2}\boldsymbol{\beta}_{h2} + \mathbf{e}_{ht2} = \begin{bmatrix} 0 \\ \boldsymbol{\gamma}_{h2} \end{bmatrix} + \mathbf{X}_{cht2}\boldsymbol{\delta}_{h2} + \mathbf{e}_{ht2}, \tag{8}$$

where $\mathbf{X}_{cht1}$ and $\mathbf{X}_{cht2}$ are matrices containing the causal variables and $\boldsymbol{\gamma}_{h1}$ and $\boldsymbol{\gamma}_{h2}$ are brand-specific constants for Category-1 and Category-2 respectively. For identification purposes, the first brand-specific constant in each category is forced to zero. Note that the parameters across the two categories, $\boldsymbol{\gamma}_{h1}$, $\boldsymbol{\gamma}_{h2}$, $\boldsymbol{\delta}_{h1}$ and $\boldsymbol{\delta}_{h2}$ are correlated.

While complete information is available for reference households, only choice information is available for the target households. Therefore, for the target group we can only specify a cross-category intercept-level model, i.e.,

$$\mathbf{u}_{ht1} = \begin{bmatrix} 0 \\ \boldsymbol{\mu}_{h1} \end{bmatrix} + \mathbf{e}_{ht1},$$
$$\mathbf{u}_{ht2} = \begin{bmatrix} 0 \\ \boldsymbol{\mu}_{h2} \end{bmatrix} + \mathbf{e}_{ht2}, \tag{9}$$

where $\boldsymbol{\mu}_{h1}$ and $\boldsymbol{\mu}_{h2}$ are assumed to be jointly distributed multivariate normal across households and, therefore, capture the underlying cross-category correlation.

To leverage information from the reference group analysis appropriately, we first need to map the posterior distribution for $\boldsymbol{\mu}_h = \{\boldsymbol{\mu}_{h1}, \boldsymbol{\mu}_{h2}\}$ into a distribution for the full set of parameters $\boldsymbol{\beta}_h = \{\boldsymbol{\beta}_{h1}, \boldsymbol{\beta}_{h2}\}$. The identification restrictions in the cross-category model with complete information are different from those required for the intercepts-only model for the target households. In particular, the first intercept is set to zero in both categories in the intercept-only model and this necessitates taking utility differences in the complete model so as to make the parameters across the two models comparable (see Rossi et al., 1996 for an extended discussion). After making adjustments for these differences in identification conditions for the above two sets of equations, and after taking iterated expectations $E[\mathbf{u}] = E_x[E[\mathbf{u}|\mathbf{X}]] = E_x[\mathbf{X}\boldsymbol{\beta}]$ we arrive at the following equalities for the parameters within the two models

$$\boldsymbol{\gamma}_{h1} + \mathbf{R}_1\boldsymbol{\delta}_{h1} = \boldsymbol{\mu}_{h1},$$
$$\boldsymbol{\gamma}_{h2} + \mathbf{R}_2\boldsymbol{\delta}_{h2} = \boldsymbol{\mu}_{h2}, \tag{10}$$

where,

$$\mathbf{R}_1 = \begin{bmatrix} \overline{x_2'} - \overline{x_1'} \\ \overline{x_3'} - \overline{x_1'} \\ \vdots \\ \overline{x_n'} - \overline{x_1'} \end{bmatrix}$$

and $\overline{x_j'}$ is a vector of the means for the causal variables of the jth brand in Category-1. The matrix $\mathbf{R}_2$ is defined analogously for Category-2.

These two sources of information can be combined using the following matrix equality

$$\begin{bmatrix} \boldsymbol{\mu}_h \\ \boldsymbol{\delta}_h \end{bmatrix} = \begin{bmatrix} \mathbf{A}_{11} & \mathbf{A}_{12} \\ \mathbf{A}_{21} & \mathbf{A}_{22} \end{bmatrix} \begin{bmatrix} \boldsymbol{\beta}_{h1} \\ \boldsymbol{\beta}_{h2} \end{bmatrix}, \tag{11}$$

where, $\boldsymbol{\mu}_h \equiv (\boldsymbol{\mu}_{h1}', \boldsymbol{\mu}_{h2}')'$, $\boldsymbol{\delta}_h \equiv (\boldsymbol{\delta}_{h1}', \boldsymbol{\delta}_{h2}')'$, $\boldsymbol{\beta}_{h1} \equiv (\boldsymbol{\gamma}_{h1}, \boldsymbol{\delta}_{h1})'$, $\boldsymbol{\beta}_{h2} \equiv (\boldsymbol{\gamma}_{h2}, \boldsymbol{\delta}_{h2})'$ and,

$$\begin{aligned} \mathbf{A}_{11} &= \begin{bmatrix} \mathbf{I} & \mathbf{R}_1 \\ \mathbf{0} & \mathbf{0} \end{bmatrix}, \quad \mathbf{A}_{12} = \begin{bmatrix} \mathbf{0} & \mathbf{0} \\ \mathbf{I} & \mathbf{R}_2 \end{bmatrix}, \\ \mathbf{A}_{21} &= \begin{bmatrix} \mathbf{0} & \mathbf{I} \\ \mathbf{0} & \mathbf{0} \end{bmatrix}, \quad \mathbf{A}_{22} = \begin{bmatrix} \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \mathbf{I} \end{bmatrix}. \end{aligned} \tag{12}$$

The reference group provides estimates of the population parameters $\boldsymbol{\alpha}$ and $\boldsymbol{\Lambda}$. These estimates allow us to write the following

$$\begin{bmatrix} \boldsymbol{\mu}_h \\ \boldsymbol{\delta}_h \end{bmatrix} \sim N(\mathbf{M}, \mathbf{Q}), \tag{13}$$

where,

$$\mathbf{M} = \begin{bmatrix} \mathbf{A}_{11}\mathbf{Z}_{h1}\boldsymbol{\alpha}_1 + \mathbf{A}_{12}\mathbf{Z}_{h2}\boldsymbol{\alpha}_2 \\ \mathbf{A}_{21}\mathbf{Z}_{h1}\boldsymbol{\alpha}_1 + \mathbf{A}_{22}\mathbf{Z}_{h2}\boldsymbol{\alpha}_2 \end{bmatrix} \tag{14}$$

and

$$\mathbf{Q} = \begin{bmatrix} \mathbf{Q}_{11} & \mathbf{Q}_{12} \\ \mathbf{Q}_{21} & \mathbf{Q}_{22} \end{bmatrix}. \tag{15}$$

In the above equation,

$$\begin{aligned}\mathbf{Q}_{11} &= \mathbf{A}_{11}^2\mathbf{\Lambda}_{11} + \mathbf{A}_{12}^2\mathbf{\Lambda}_{22} + \mathbf{A}_{11}\mathbf{A}_{12}(\mathbf{\Lambda}_{12} + \mathbf{\Lambda}_{21}),\\ \mathbf{Q}_{22} &= \mathbf{A}_{21}^2\mathbf{\Lambda}_{11} + \mathbf{A}_{22}^2\mathbf{\Lambda}_{22} + \mathbf{A}_{21}\mathbf{A}_{22}(\mathbf{\Lambda}_{12} + \mathbf{\Lambda}_{21}),\\ \mathbf{Q}_{12} &= \mathbf{A}_{11}\mathbf{A}_{21}\mathbf{\Lambda}_{11} + \mathbf{A}_{12}\mathbf{A}_{22}\mathbf{\Lambda}_{22} + \mathbf{A}_{11}\mathbf{A}_{22}\mathbf{\Lambda}_{12} + \mathbf{A}_{12}\mathbf{A}_{21}\mathbf{\Lambda}_{21},\end{aligned} \tag{16}$$

and

$$\mathbf{Q}_{21} = \mathbf{Q}_{12}'.$$

Given the above setup, the conditional distribution for the causal coefficients $\boldsymbol{\delta}_h$, is a normal $N(\boldsymbol{\delta}_h|\boldsymbol{\mu}_h)$ and can be written as:

$$\begin{aligned}\boldsymbol{\delta}_h|\boldsymbol{\mu}_h &\sim N(E(\boldsymbol{\delta}_h|\boldsymbol{\mu}_h), \mathrm{Var}(\boldsymbol{\delta}_h|\boldsymbol{\mu}_h)),\\ E(\boldsymbol{\delta}_h|\boldsymbol{\mu}_h) &= \mathbf{A}_{21}\mathbf{Z}_{h1}\boldsymbol{\alpha}_1 + \mathbf{A}_{22}\mathbf{Z}_{h2}\boldsymbol{\alpha}_2 + \mathbf{Q}_{21}\mathbf{Q}_{11}^{-1}(\boldsymbol{\mu}_h - (\mathbf{A}_{11}\mathbf{Z}_{h1}\boldsymbol{\alpha}_1 + \mathbf{A}_{12}\mathbf{Z}_{h2}\boldsymbol{\alpha}_2)),\\ \mathrm{Var}(\boldsymbol{\delta}_h|\boldsymbol{\mu}_h) &= \mathbf{Q}_{22} - \mathbf{Q}_{21}\mathbf{Q}_{11}^{-1}\mathbf{Q}_{12}.\end{aligned} \tag{17}$$

It can be seen from the above expressions that $\mathbf{\Lambda}_{12}$, the cross-category covariance between the parameters, contributes to both the mean and variance of the conditional distribution for $\boldsymbol{\delta}_h$.

4. Simulation study

At this point, we could simply present an empirical application to illustrate how our model works (we provide such an application in the next section). However, it is difficult to draw broader and generalizable results from any single application. Even analyzing multiple data sets has limitations since (a) we do not know the ‘‘truth’’ that generated these data and (b) the underlying factors that make these data different are not varied systematically. For example, industry reports suggest that cross-selling works in some cases but not in others. By simply examining these cases, it is not obvious under what conditions cross-selling works. Therefore, to gain a better understanding of the conditions that are necessary for transfer of information across categories, we conduct a detailed simulation.

4.1. *Factors*

Based on the literature and our understanding of the phenomena, we chose three key factors for generating the simulation data sets. These factors and their levels are briefly described below.

1. *Correlation in response parameters* (ρ)*:* As mentioned in our model section, if the correlation among the parameters of the two categories Λ_{12} is zero, then there is no information transfer from one category to the other. Clearly, the higher this correlation, the greater is the information leveraged. Although we allow all parameters to be correlated, we focus on the correlation between the price parameter of the two product categories. Previous studies have shown varying levels of this correlation. For example, Ainslie and Rossi (1998) examined five product categories and found the mean correlation of 0.28 in the price sensitivities across these categories. Erdem et al. (2001) investigated three product categories and found correlation among price coefficients in the range of 0.56 to 0.68. Therefore, for our simulation study we chose two levels, 0.3 to represent a relatively low level of correlation, and 0.7 to represent a relatively high level of correlation among parameters.
2. *Consumer heterogeneity* (γ)*:* In addition to parameter correlation, consumer heterogeneity is also likely to play a significant role in information leverage across two categories. Recall that we have a reference set of households for whom we have complete information in both categories. Our objective is to make inferences about the target households for whom we have limited information in one or both categories. If consumers are very homogeneous (in the extreme case, identical), then information on reference households within a category is sufficient to make inferences about the behavior of target households in that category. In other words, additional information from a second category, no matter how high the parameter correlation across these two categories may be, is unlikely to provide any additional information. Although heterogeneity estimates from previous studies are not strictly comparable due to differences in models and other features, nonetheless they provide guidelines for our choice in this simulation. In their study of tuna, Rossi et al. (1996) find unobserved heterogeneity in price parameter (measured as variance of price parameter across consumers) to be approximately 11. Ainslie and Rossi (1998) found the comparable heterogeneity in price parameter across five product categories in the range of 4.88 to 9.0. Erdem et al. (2001) estimated this heterogeneity in three categories in the range of 0.37 to 1.69. Based on these studies, we chose two levels of consumer heterogeneity (defined as variance of parameters across consumers) as 0.3 (low) and 11 (high).
3. *Number of observations* (n_1, n_2)*:* Bayesian theory suggests that information transfer from Category-1 to Category-2 also depends on the precision of household-level estimates in Category-1. Among other things, this precision depends on the number of observations available for a household in Category-1. The number of observations for a household may differ across categories because

of differences in purchase frequency. Based on previous scanner studies, we chose two levels for the number of observations (n_1) in the first category – 8 (low) and 24 (high). The number of observations for the second category (n_2) were kept constant at eight.

4.2. *Data generation*

We generated data using the cross-category multinomial probit model described in Section 3.1. We considered three brands in each category and constructed the responses assuming that the underlying latent variables are affected by only one explanatory variable, say, brand price. As mentioned above, we allowed the number of observations per household to differ across the two categories. For each observation, the three brand prices in a category are independently and identically distributed as a uniform random deviate between 2.00 and 5.00.[4] Thus, for observation t of household h within category k, the 3×3 design matrix, $\mathbf{X}_{htk}$, contains the two intercepts and a column of prices for the three brands. Given the design matrix, the underlying latent variables for household h in category k follow $\mathbf{u}_{htk} \sim N(\mathbf{X}_{htk}\boldsymbol{\beta}_{hk}, \boldsymbol{\Sigma}_k)$, where $\boldsymbol{\Sigma}_k$ is fixed to be identity and $\boldsymbol{\beta}_{hk}$ is a 3×1 coefficient vector.

The household-level parameters, $\boldsymbol{\beta}_{h1}$ and $\boldsymbol{\beta}_{h2}$, are drawn from a multivariate normal with a population mean $\boldsymbol{\mu}$ of dimension 6×1 and a population variance-covariance matrix $\boldsymbol{\Lambda}$ of dimension 6×6. We did not consider any household-level demographics for this simulation, thus $\mathbf{Z}_h = \mathbf{I}$. We fixed $\boldsymbol{\mu}'$ at $[0.9, 1.5, -1.5, 1.0, 0.8, -2.0]$ for all data sets in the simulation and set Λ to be an equicorrelated matrix of the form $\gamma\boldsymbol{\Omega}$. The correlation matrix $\boldsymbol{\Omega}$ has a correlation ρ on each off-diagonal element and γ represents the population variance for each parameter.

As the three factors correlation, heterogeneity, and number of observations are varied at two levels each, we have a $2 \times 2 \times 2$ design for the simulation study. Within each of the eight cells, we generated 10 replications involving a reference group of 200 households and a target group of 100 households. The design matrices ($\mathbf{X}_{htk}$) for these 80 replications were randomly generated using uniform deviates to create variation across the replications.

4.3. *Results*

We assess the simulation results on two criteria—likelihood of the target group's data under different information Scenarios and how well household-specific

4 A uniform distribution can increase the variance in prices more than what is observed in scanner data. We thank one of the reviewers for bringing this to our attention.

parameters are recovered. We use the error sum of squares (ESS) to measure how well household-specific parameters are estimated for the target group. The ESS is defined by the following formula:

$$\mathrm{ESS} = \frac{1}{I} \sum_{i=1}^{I} \sum_{j=1}^{3} (\beta_{ij} - \hat{\beta}_{ij})^2, \tag{18}$$

where β_{ij} is the true (known) parameter j for consumer i within the target category, $\hat{\beta}_{ij}$ is the estimated posterior mean for the parameter, and I is the number of individuals in the target group.

For computing the data likelihood, we use the estimated household-specific posterior mean of the parameters, $\hat{\beta}_i$, as an input to the Geweke–Keane–Hajivassiliou (GHK) algorithm (Geweke, 1991; Hajivassiliou, 1990; Keane, 1994) for the multinomial probit model. The algorithm details can be found in Chib et al. (1998).

We compute ESS and the data likelihood for target households in a category using both single-category and cross-category analysis. We then compare the improvement in these two criteria in going from single-category analysis to cross-category analysis under different experimental conditions. In other words, our focus is to see the incremental benefit of the cross-category analysis that leverages information from both categories over a single-category analysis that treats each category as independent of the other. Without loss of generality, we assume that we are leveraging the information from Category-1 to predict consumer behavior in Category-2. We now discuss the results of four key Scenarios (1, 4, 5 and 7). Recall that these Scenarios represent varying degrees of information about target households. In every Scenario, we have complete information about reference households in both categories.

4.3.1. Scenario-1: Base-Base. This Scenario represents the situation where a firm has no specific information about a target customer. The firm, however, does know the distribution of preferences and price sensitivities across reference consumers in both product categories, as well as the correlation in these preferences and price sensitivities between the two products. In a single-category analysis, inference for the target household will be simply based on a random draw from the distribution of reference households in Category-2 (Rossi et al., 1996). This analysis can be possibly improved by using information about households' preference distribution in Category-1 and its correlation with Category-2. We call the latter as cross-category analysis. Table 2 presents the absolute improvement in log-likelihood and the percent improvement in ESS in going from single-category to cross-category analysis, averaged over the 10 replicated data sets within each experimental cell of the $2 \times 2 \times 2$ design. Here, each of the three factors (correlation in parameters between the two products, consumer heterogeneity in parameters, and number of

Table 2. Simulation results for Scenario-1: Base-Base.

Parameter correlation	Consumer heterogeneity	Number of observations	Percent improvement in ESS	Improvement in LL
L	L	L	− 0.53	− 5.18
L	L	H	0.52	− 2.60
L	H	L	0.53	3.93
L	H	H	0.12	− 5.86
H	L	L	− 0.14	− 1.76
H	L	H	0.41	− 1.18
H	H	L	0.03	− 18.19
H	H	H	1.33	8.50

observations in Category-1) are represented at two levels (L = low, and H = high). The low and high levels for each factor are as discussed above. Statistically significant improvements in log-likelihood and ESS are shown in bold.

The results show that cross-category analysis does not provide any significant improvement in either the data likelihood or parameter recovery (as measured by ESS) for any of the experimental conditions. In other words, additional information from the second category (e.g., correlation in price sensitivities, even when it is as high as 0.7), does not provide any incremental benefits. This is not surprising since the base case does not provide any specific information about the target households. This result sets a benchmark for the potential improvement in other Scenarios.

4.3.2. Scenario-4: Full-Base/Demo. As mentioned before, this Scenario represents a typical cross-selling case. Here a firm may have full information about target customers in one category (Category-1) but no, or only demographic, information in Category-2. The firm wishes to leverage customers' information in Category-1 to predict their behavior in Category-2. For simplicity, we ignore demographics in this simulation. To ensure consistency with the nomenclature in Table 1, we continue to call this as Scenario-4.

To compute the incremental benefit we compare the predictions in Category-2 from a single-category model with those from our cross-category model. In a single-category probit model, we have no specific information about target households in Category-2. Therefore, inferences can only be made by using the population-level estimates obtained from estimating the model on the reference group purchases in Category-2. In contrast, as described in Section 3.2, when a cross-category model is used, inferences about the target group within Category-2 are based on (a) the population-level estimates obtained from estimating this joint model on the reference group purchases in both categories and (b) the full purchase history of the target households in Category-1. A comparison of the results from the two approaches provides an indication of the incremental information gleaned from a cross-category analysis. Table 3 provides these results. The numbers in bold represent a significant improvement.

Table 3. Simulation results for Scenario-4: Full-Base.

Parameter correlation	Consumer heterogeneity	Number of observations	Percent improvement in ESS	Improvement in LL
L	L	L	− 5.79	− 7.73
L	L	H	1.78	0.85
L	H	L	**9.89**	4.75
L	H	H	**11.88**	4.62
H	L	L	**21.13**	2.45
H	L	H	**37.61**	5.25
H	H	L	**38.41**	**117.23**
H	H	H	**51.38**	**114.02**

We begin by focusing on log-likelihood. Our results show that cross-category information improves the data likelihood only when both parameter correlation and consumer heterogeneity are high. If these conditions hold, data likelihood from the cross-category model improves over those from the single-category analysis by as much as 115 points. In all other conditions, there is no significant improvement. It is reasonable to expect that when correlation among parameters of the two categories is low ($\rho = 0.3$), there is very little transfer of information from one category to the other. Therefore, it is not surprising to find that high parameter correlation is a necessary condition for improvement in prediction over single category analysis. However, a somewhat surprising result is that high correlation, while necessary, is not a sufficient condition. Why? When consumer heterogeneity is low (in the extreme case consumers are perfectly homogeneous), information about reference consumers in Category-2 is a strong indicator of target consumers' behavior in this category as consumers are relatively homogeneous. Additional information from Category-1 purchases of target customers does not provide any incremental benefit in predicting their behavior. Therefore, high heterogeneity is needed in addition to high correlation.

We next focus on ESS. The ESS results show that in all conditions, except when both heterogeneity and correlation are low, the full purchase information of a target household in Category-1 is useful in recovering its parameters in Category-2. Further, the incremental benefit of this information is the highest under conditions of high heterogeneity and high correlation. When all three factors are at a high level, cross-category model ESS improves over single-category ESS by as much as 51%. Finally, we rank the factors in terms of their influence on the results by averaging the gain for the three factors and considering the difference in the gain between the high and low level of each factor. For this analysis, the non-significant gains are taken to be 0. The average gain for high level of correlation is 37% whereas that for a low level of correlation is 11%. This gives a spread of 26%. The spread for heterogeneity and number of observations can be similarly computed and they are − 2% and 10% respectively. Thus, correlation is the most influential factor, followed by number of observations and consumer heterogeneity.

Comparing the results for log-likelihood and ESS, we see that in several situations while ESS improves with additional information from a second category, there is no significant improvement in log-likelihood. In other words, the benefit of additional information from a second category depends on the researcher's objective.

4.3.3. Scenario-5: Choice-Choice. This Scenario represents situations where companies collect information about consumers' choices but do not have any information about the causal variables (e.g., price) that may affect their choices. Examples include direct marketing companies that track consumers' choices simply because they have to ship the product. However, many of these companies do not keep records of all the events and promotional material sent to every customer. Similarly, loyalty programs keep detailed account of customers' purchases (e.g., hotel stay, travel etc.) but not necessarily the price paid or other causal variables.

In our context, we once again have complete information for both product categories for the reference households, but only choice information in both categories for the target households. For a single category, Rossi et al. (1996) show how we can draw inferences about the price sensitivity of target households without observing causal information for them. In our context of two categories, we wish to find out if correlation between these two categories and the choice information of target households in Category-1 help us draw better inferences about their preferences and purchase behavior in Category-2.

Once again we compare the predictions for target households in Category-2 from a single-category model with those from our cross-category model. In a single-category model, inferences about the parameters for the target group can be made by combining (a) the information about choices of the target group in Category-2 and (b) the population-level estimates obtained from estimating this model on the reference group purchases in Category-2. When a cross-category model is used, inferences about the parameters for the target group in Category-2 are based on (a) choice information of the target households in Category-2 (b) the population-level estimates obtained from estimating this joint model on the reference group purchases in both categories (which includes the correlation in parameters between these two categories) and (c) information about choices of the target households in Category-1. A comparison of the results from the two approaches provides an indication of the incremental benefit from the cross-category analysis.

Table 4 provides these results. As earlier, this table shows the absolute improvement in log-likelihood and the percent improvement in ESS of cross-category model over single category model. The improvement estimates are averaged over the 10 replicated data sets within each combination of the experimental factors. Statistically significant changes are highlighted in bold.

We begin with the log-likelihood results. In general, there is no improvement in log-likelihood. In fact, under low correlation and low heterogeneity, there is a significant decline! We explain this surprising finding in the following discussion of the ESS results.

Table 4. Simulation results for Scenario-5: Choice-Choice.

Parameter correlation	Consumer heterogeneity	Number of observations	Percent improvement in ESS	Improvement in LL
L	L	L	**− 5.38**	**− 22.72**
L	L	H	**− 7.95**	**− 17.60**
L	H	L	3.51	− 35.15
L	H	H	0.70	− 23.87
H	L	L	− 4.44	− 3.75
H	L	H	**9.72**	− 1.50
H	H	L	1.33	− 43.45
H	H	H	1.18	− 23.95

The ESS results show that under low correlation and low heterogeneity condition, the cross-category model performs worse than the single category model. In other words, instead of being beneficial, additional information is detrimental to parameter recovery! We believe this happens for the following reasons. In a choice-choice context where we have no information about causal variables for target households, it is obviously much harder to infer price sensitivity for these customers. In case of low heterogeneity, customers are very similar. Therefore, price sensitivity of reference households in Category-2 provides significant information about the price sensitivity of target households in Category-2. In effect, this is what a single-category model is doing. When we add cross-category information in a low correlation context, we are doing the following. First, using choice information of target households in Category-1, we are inferring their price sensitivity in Category-1. Next, we are using the parameter correlation between the two categories (estimated from reference households) to link Category-1 price sensitivity of a target household to its price sensitivity in Category-2. Note, in a low heterogeneity case, analyzing only Category-2 provides us a fairly accurate account of target household's price sensitivity. In the extreme case of completely homogeneous population, we have a perfect estimate of target households' price sensitivity in a single category analysis. However, when we add the noisy (due to low correlation) estimates from the cross-category analysis to good (in the extreme case, perfect) estimates of single-category analysis, we end up with estimates that are worse than single category estimates. The inaccuracy in the estimates in then reflected in the deterioration of the log-likelihood. Most of the other differences in ESS are insignificant except H-L-H combination of correlation, heterogeneity and observations respectively. However, this appears to be an anomaly.

4.3.4. Scenario-7: Full-Full. This is the most data rich Scenario where a firm has complete information about both reference and target households in both product categories. Most cross-category studies (e.g., Ainslie and Rossi, 1998; Erdem et al., 2001) consider this Scenario. Results for this Scenario are given in Table 5.

Table 5. Simulation results for Scenario-7: Full-Full.

Parameter correlation	Consumer heterogeneity	Number of observations	Percent improvement in ESS	Improvement in LL
L	L	L	1.14	− 0.08
L	L	H	5.63	− 4.9
L	H	L	5.36	0.14
L	H	H	3.61	− 0.49
H	L	L	**17.76**	**5.70**
H	L	H	**28.71**	1.38
H	H	L	**15.46**	− 1.34
H	H	H	**15.37**	− 4.96

These results suggest that compared to a single-category analysis, cross-category analysis in general does not improve the data likelihood. The H-L-L combination of correlation, heterogeneity and observations appears to be an anomaly. However, there is a significant improvement in parameter recovery or ESS under high correlation case. We can rank the three factors in terms of their influence on the ESS results. We average the gain for the three factors and consider the difference in the gain between the high and low level of each factor. The average gain under high correlation is 19.32% and under low correlation is 0% (non-significant gains are taken to be 0). Thus, the spread in gain is 19.32%. The differences for consumer heterogeneity and number of observations are smaller at − 7.82% and 5.43% respectively. Thus, correlation is the most influential factor followed by consumer heterogeneity and number of observations.

4.4. *Summary of simulation results*

Based on our simulation results, we can draw the following conclusions about the benefit of cross-category analysis over single-category analysis.

1. In most situations there is very little gain in data likelihood compared to single category analysis. This suggests that in any empirical application, where the likelihood provides a way to compare approaches, a cross-category analysis is unlikely to show any significant gains over a single-category analysis. This makes the single-category analysis as suggested by Rossi et al. (1996) even more powerful than we may have realized in the past. In many situations, however, leveraging information across categories improves the recovery of underlying household-specific parameters. This finding is managerially useful as better estimation of individual-level parameters can lead to more optimal customization of marketing initiatives.
2. Cross-category analysis and leveraging information across categories is most useful when there is complete information for a customer in one category and no

or very little information about that customer in a related category (Scenario 4). However, when firms have no causal information in either product category (Scenario 5), they may be well advised to avoid the complexity of cross-category analysis since it does not provide any benefit over single-category analysis.

3. In many situations, high correlation among parameters of two categories is a necessary but not sufficient condition for transfer of information between two products. Consumer heterogeneity also plays an important part. This provides strong guidelines for researchers and managers about when additional information and complexity is likely to pay off.
4. More information is not always better. In some cases (low correlation, low heterogeneity and choice only), including additional information from a second category may lead to worse results.

5. Application

In this section, we report the results from an application for two grocery products. Here, we use our model on the reference households to get estimates of parameter correlations and consumer heterogeneity. In addition, the data provide us information about the number of observations in each category. The information on these three factors when combined with our simulation guidelines allows us to predict how well a cross-category model is likely to perform compared to a single-category model. In other words, we can hypothesize the value of leveraging information across categories. We estimate both single- and cross-category models, compare the likelihood and assess whether our application results confirm our simulation guidelines.

5.1. Data

The data, made available by A.C. Nielsen, pertains to two related product categories—Breakfast Foods and Table Syrup. We deliberately chose related categories to ensure significant correlation in parameters between these two products. The data span a period of 120 weeks from January 1993 to March 1995 and are from a large metropolitan market area in the western United States. We randomly chose 200 people to form the reference group and 100 people to form the target group. There are four major brands within each category. For each brand, we have price and promotion information. Promotion is a dummy variable created by combining various promotional vehicles such as feature and display. On average, we have eight observations per household in Breakfast Foods and three observations per household in Table Syrup. Table 6 provides summary statistics for the two categories. The column labeled promotion indicates the proportion of times that a specific brand was promoted.

Table 6. Summary statistics of data.

Breakfast Foods			Table Syrup		
Brand	Price ($)	Promotion	Brand	Price ($)	Promotion
1	1.75	0.07	1	1.98	0.03
2	1.58	0.04	2	2.87	0.02
3	1.92	0.09	3	1.74	0.05
4	1.94	0.01	4	1.63	0.07

5.2. *Estimation*

Both the single and cross-category models were estimated using the Bayesian procedures outlined in Appendix A. We first estimated the population parameters using data for reference customers. The population parameters and the purchase information of target households were then used in tandem to estimate the parameters for the target households and to predict their behavior under the four information Scenarios discussed earlier. The MCMC algorithm for estimating the population level parameters in both single category and cross-category analysis was run for 40,000 iterations with a burn-in period of 25,000 iterations. Convergence of the parameters was assessed by monitoring the time series plots across the MCMC iterations.

5.3. *Results*

5.3.1. Results for reference customers. The parameter estimates from the independent single-category analysis of reference households are given in Table 7. This table shows the estimates for the population mean and the population variance

Table 7. Parameter estimates for single-category analysis.

	Breakfast Foods		Table Syrup	
Parameter	Mean	Variance	Mean	Variance
Intercept1	**− 0.67**	3.90	− 0.02	1.96
	(−1.11, −0.23)	(2.48, 5.73)	(−0.75, 0.63)	(0.74, 4.14)
Intercept2	**2.08**	2.94	− 0.54	3.42
	(1.69, 2.57)	(1.72, 4.80)	(−1.13, 0.28)	(1.83, 5.94)
Intercept3	**−0.94**	4.30	**− 1.45**	1.67
	(−2.18, −0.19)	(2.06, 8.79)	(−2.03, −0.86)	(0.67, 3.35)
Price	**− 3.36**	2.89	**− 5.18**	1.48
	(−4.20, −2.88)	(0.93, 4.44)	(−5.91, −4.62)	(0.37, 3.07)
Promotion	0.28	1.40	**1.22**	1.71
	(−0.17, 0.74)	(0.63, 2.78)	(0.29, 2.10)	(0.52, 4.25)

of the parameters in the two categories. The numbers in parentheses show the 95% posterior intervals around the estimate and "significant" posterior means are shown in bold. The corresponding estimates from a cross-category analysis are given in Table 8. In addition, the cross-category analysis also provides estimates of the correlation between parameters of the two categories (Table 9). Note, in our estimation procedure, we allow all the parameters across the two categories to be correlated with each other. In other words, we do not impose any structure on the $\boldsymbol{\Lambda}$ matrix. We estimate the $\boldsymbol{\Lambda}$ matrix during each iteration of the MCMC algorithm and then calculate the correlations. Table 9 gives the average correlations across 25,000 iterations with the significant ones shown in bold. Significance is ascertained by looking at the 95% confidence interval around the estimate.

From these tables we make the following observations. First, the parameter estimates have face validity, i.e., the price parameters are negative and the promotion parameters have positive signs. Second, the parameter estimates from the single-category analysis are very similar to those obtained from the cross-category analysis. Third, while there is no significant correlation between the promotion parameters of the two categories, the price parameters are positively correlated. However, this correlation is only 0.21 which is consistent with the results in Ainslie and Rossi (1998). Fourth, consumer heterogeneity, as measured by the population variances, is between 1.7 to 4.3. In general, there is greater heterogeneity in Breakfast Foods than in Table Syrup.

Recall that our simulation results were based on three factors: parameter correlation (low = 0.3, high = 0.7), consumer heterogeneity (low = 0.3, high = 11) and number of observations (low = 8, high = 24). The results from our application show low correlation between parameters (0.21 to − 0.32), low to moderate customer heterogeneity (1.7 to 4.3), and low number of observations (3 for Table Syrup and 8 for Breakfast Foods). Therefore, our simulation guidelines suggest that cross-category analysis is unlikely to provide a better data likelihood than single-category

Table 8. Parameter estimates for cross-category analysis.

	Breakfast Foods		Table Syrup	
Parameter	Mean	Variance	Mean	Variance
Intercept1	**−0.64**	4.29	− 0.22	2.62
	(−1.08, −0.19)	(2.67, 6.38)	(−0.97, 0.52)	(1.11, 4.99)
Intercept2	**2.13**	3.33	− 0.63	4.36
	(1.74, 2.59)	(1.96, 5.41)	(−1.28, 0.03)	(1.84, 8.51)
Intercept3	**−0.80**	4.32	**−1.68**	2.45
	(−1.63, −0.12)	(2.27, 7.58)	(−2.35, −1.04)	(1.06, 4.92)
Price	**−3.38**	3.70	**−5.81**	1.89
	(−3.90, −2.91)	(2.02, 5.90)	(−6.75, −4.80)	(0.83, 3.24)
Promotion	0.29	1.73	**1.53**	2.45
	(−0.15, 0.75)	(0.83, 3.23)	(0.47, 2.67)	(0.90, 5.59)

Table 9. Cross-category correlation matrix.

	Category-2				
Category-1	Intercept-1	Intercept-2	Intercept-3	Price	Promotion
Intercept-1	0.23	− 0.19	**−0.32**	0.23	0.01
Intercept-2		**−0.27**	− 0.10	0.21	0.02
Intercept-3			− 0.28	− 0.18	0.03
Price				**0.21**	0.03
Promotion					0.04

analysis for all possible information Scenarios. While we do not know the true parameters in the context of our application, simulation results suggest that parameter recovery is also unlikely to benefit in almost all Scenarios, with the possible exception of Scenario-4 (full-base).

5.3.2. ***Results for target customers.*** In each of the four information Scenarios discussed earlier, we use both the single- and cross-category models to estimate parameters for the target households. In the single-category analysis for each product, we use full information of the reference households and any available information (depending on the Scenario) of the target households in that product category only. In other words, we completely ignore the purchase information in the related product category. For example, to estimate parameters of target households in Breakfast Foods in Scenario-5 (choice-choice), we use full information of reference households but only choice information of target households in Breakfast Foods. Similarly, in Scenario-4 (full-base), we have no specific purchase information on target households. In cross-category analysis, we use full information of reference households in both categories and any available information of target households in both categories. For example, cross-category analysis for Breakfast Foods in Scenario-4 (full-base) involves using full information of the reference households in both the categories, full information about the target households in Table Syrup, but no specific information about target households in Breakfast Foods.

Using the single-category and cross-category analysis, we estimated the mean price parameter for each of the 100 target households. We then calculated the sample mean and standard deviations for these household-specific parameters. Table 10 shows the mean price parameter estimates with the standard deviation in the parenthesis. As expected from our simulation results, there are no significant differences in the price parameter estimates of single- and cross-category analysis for any Scenario, except for Scenario-4 (the results are similar for other parameters).

We also computed the log-likelihood using both the single- and cross-category models. As expected from our simulation results, cross-category analysis did not lead to any significant improvement in likelihood over a single category analysis for any of the four scenarios.

Table 10. Mean price parameter estimates.

Information scenarios	Breakfast Foods: Single category analysis	Breakfast Foods: Cross- category analysis	Table Syrup: Single category analysis	Table Syrup: Cross-category analysis
Scenario-1	– 3.37	– 3.38	– 5.19	– 5.81
(Base-base)	(0.04)	(0.02)	(0.31)	(0.40)
Scenario-4	**–3.37**	**–5.26**	**–5.19**	**–3.51**
(Full-base/demo)	(0.04)	(0.39)	(0.31)	(0.37)
Scenario-5	– 3.23	– 2.74	– 3.17	– 3.34
(Choice-choice)	(1.58)	(1.24)	(0.43)	(0.35)
Scenario-7	– 3.38	– 3.40	– 5.65	– 5.83
(Full-full)	(1.33)	(1.50)	(0.30)	(0.45)

The number in parentheses is the standard deviation. Significant differences are shown in bold.

6. Extension to multiple categories

Until now the paper has focused on two categories. Extension to multiple categories is straightforward. Here we show the extension of our model for three categories. The number of information scenarios (Table 1) grows if an additional category is considered. In what follows, for illustration purposes, we choose two information Scenarios and show how inferences for household-level parameters can be carried out. The two Scenarios that we choose are Full-Choice-Demographics (Full information in Category-1, Choice information in Category-2 and Only Demographics in Category-3) and Full-Full-Choice (Full information in Category-1, Full information in Category-2 and Choice information in Category-3).

6.1. Full-Choice-Demographics

In this Scenario, we have full information about the target group in Category-1, choice information in Category-2 and only demographics in Category-3. The following analysis shows how to combine these three information sets to infer household-specific parameters in categories where only partial information is available.

We begin with Category-2. The information set in that category is choice only. Specifically, utility for a complete data model in Category-2 (given in equation (1)) can be rewritten as

$$\mathbf{u}_{ht2} = \mathbf{X}_{ht2}\boldsymbol{\beta}_{h2} + \mathbf{e}_{ht2} = \begin{bmatrix} 0 \\ \boldsymbol{\gamma}_{h2} \end{bmatrix} + \mathbf{X}_{cht2}\boldsymbol{\delta}_{h2} + \mathbf{e}_{ht2}, \tag{19}$$

where $\mathbf{X}_{cht2}$ is a matrix containing the causal variables and $\boldsymbol{\gamma}_{h2}$ is brand-specific constants for Category-2. For identification purposes, the first brand-specific constant is forced to zero.

While complete information is available for reference households, only choice information is available for the target households. Therefore, for the target group we can only specify an intercept-level model, i.e.,

$$\mathbf{u}_{ht2} = \begin{bmatrix} 0 \\ \boldsymbol{\mu}_{h2} \end{bmatrix} + \mathbf{e}_{ht2}. \tag{20}$$

To leverage information from the reference group analysis appropriately, we need to map the posterior distribution for $\boldsymbol{\mu}_{h2}$ into a distribution for the full set of parameters $\boldsymbol{\beta}_{h2}$. The identification restrictions in the model with complete information are different from those required for the intercepts-only model for the target households. In particular, the first intercept is set to zero in both categories in the intercept-only model and this necessitates taking utility differences in the complete model so as to make the parameters across the two models comparable (see Rossi et al., 1996 for an extended discussion). After making adjustments for these differences in identification conditions for the above two sets of equations, and after taking iterated expectations $E[\mathbf{u}] = E_x[E[\mathbf{u}|\mathbf{X}]] = E_x[\mathbf{X}\boldsymbol{\beta}]$ we arrive at the following equalities for the parameters within the model.

$$\gamma_{h2} + \mathbf{R}_2\boldsymbol{\delta}_{h2} = \boldsymbol{\mu}_{h2}, \tag{21}$$

where,

$$\mathbf{R}_2 = \begin{bmatrix} \overline{x_2'} - \overline{x_1'} \\ \overline{x_3'} - \overline{x_1'} \\ \vdots \\ \overline{x_n'} - \overline{x_1'} \end{bmatrix}$$

and $\overline{x_j'}$ is a vector of the means for the causal variables of the jth brand in Category-2.

We can now integrate the information in Category-2 with that in Category-1 and Category-3. This linkage explictly captures the cross-category influence. The integration is best expressed through the following matrix equality.

$$\begin{bmatrix} \boldsymbol{\beta}_{h1} \\ \boldsymbol{\mu}_{h2} \\ \boldsymbol{\delta}_{h2} \\ \boldsymbol{\beta}_{h3} \end{bmatrix} = \begin{bmatrix} \mathbf{I} & \mathbf{0} & \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \mathbf{I} & \mathbf{R}_2 & \mathbf{0} \\ \mathbf{0} & \mathbf{0} & \mathbf{I} & \mathbf{0} \\ \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{I} \end{bmatrix} \begin{bmatrix} \boldsymbol{\beta}_{h1} \\ \boldsymbol{\gamma}_{h2} \\ \boldsymbol{\delta}_{h2} \\ \boldsymbol{\beta}_{h3} \end{bmatrix}. \tag{22}$$

This can be rewritten as:

$$\begin{bmatrix}\boldsymbol{\theta}_h\\ \boldsymbol{\zeta}_h\end{bmatrix}=\begin{bmatrix}\mathbf{A}_{11} & \mathbf{A}_{12}\\ \mathbf{A}_{21} & \mathbf{A}_{22}\end{bmatrix}\begin{bmatrix}\boldsymbol{\beta}_{h12}\\ \boldsymbol{\beta}_{h3}\end{bmatrix}, \tag{23}$$

where, $\boldsymbol{\theta}_h \equiv (\boldsymbol{\beta}'_{h1}, \boldsymbol{\mu}'_{h2})'$, $\boldsymbol{\zeta}_h \equiv (\boldsymbol{\delta}'_{h2}, \boldsymbol{\beta}'_{h3})'$, $\boldsymbol{\beta}_{h12} \equiv (\boldsymbol{\beta}'_{h1}, \boldsymbol{\gamma}'_{h2}, \boldsymbol{\delta}'_{h2})'$ and,

$$\begin{aligned}\mathbf{A}_{11}&=\begin{bmatrix}\mathbf{I} & \mathbf{0} & \mathbf{0}\\ \mathbf{0} & \mathbf{I} & \mathbf{R}_2\end{bmatrix}, \quad \mathbf{A}_{12}=\begin{bmatrix}\mathbf{0}\\ \mathbf{0}\end{bmatrix},\\ \mathbf{A}_{21}&=\begin{bmatrix}\mathbf{0} & \mathbf{0} & \mathbf{I}\\ \mathbf{0} & \mathbf{0} & \mathbf{0}\end{bmatrix}, \quad \mathbf{A}_{22}=\begin{bmatrix}\mathbf{0}\\ \mathbf{I}\end{bmatrix}.\end{aligned} \tag{24}$$

The reference group provides estimates of the population distribution's fixed effect $\boldsymbol{\alpha}$, and covariance $\boldsymbol{\Lambda}$.

Using the above defined submatrices of $\mathbf{A}$, we can rewrite the multivariate normal expression as

$$\begin{bmatrix}\boldsymbol{\theta}_h\\ \boldsymbol{\zeta}_h\end{bmatrix}\sim N\left(\begin{bmatrix}\mathbf{A}_{11} & \mathbf{A}_{12}\\ \mathbf{A}_{21} & \mathbf{A}_{22}\end{bmatrix}\begin{bmatrix}\mathbf{Z}_{h12}\boldsymbol{\alpha}_{12}\\ \mathbf{Z}_{h3}\boldsymbol{\alpha}_3\end{bmatrix}, \begin{bmatrix}\mathbf{A}_{11} & \mathbf{A}_{12}\\ \mathbf{A}_{21} & \mathbf{A}_{22}\end{bmatrix}\begin{bmatrix}\boldsymbol{\Lambda}_{12} & \boldsymbol{\Lambda}_{123}\\ \boldsymbol{\Lambda}_{321} & \boldsymbol{\Lambda}_{33}\end{bmatrix}\begin{bmatrix}\mathbf{A}_{11} & \mathbf{A}_{21}\\ \mathbf{A}_{12} & \mathbf{A}_{22}\end{bmatrix}\right), \tag{25}$$

where $\boldsymbol{\Lambda}_{12}$ refers to the cross-category covariance matrix between parameters in Category-1 and Category-2, $\boldsymbol{\Lambda}_{123}$ is the cross-category covariance matrix between parameters in Category-1, Category-2 and Category-3 and $\boldsymbol{\Lambda}_{33}$ represents the within category covariance matrix of the parameters in Category-3. We choose this particular representation of the covariance matrix between the parameters of Category-1, Category-2 and Category-3 in order to have a 2×2 structure, which facilitates the subsequent analytical computation. Finally, $\mathbf{Z}_{h12}\boldsymbol{\alpha}_{12} \equiv (\mathbf{Z}_{h1}\boldsymbol{\alpha}'_1, \mathbf{Z}_{h2}\boldsymbol{\alpha}'_2)'$

For ease of illustration, we denote the covariance, $\mathbf{A}\boldsymbol{\Lambda}\mathbf{A}'$, as 2×2 matrix $\mathbf{Q}$, where,

$$\begin{aligned}\mathbf{Q}_{11}&=\mathbf{A}^2_{11}\boldsymbol{\Lambda}_{12}+\mathbf{A}^2_{12}\boldsymbol{\Lambda}_{33}+\mathbf{A}_{11}\mathbf{A}_{12}(\boldsymbol{\Lambda}_{123}+\boldsymbol{\Lambda}'_{123}),\\ \mathbf{Q}_{22}&=\mathbf{A}^2_{21}\boldsymbol{\Lambda}_{12}+\mathbf{A}^2_{22}\boldsymbol{\Lambda}_{33}+\mathbf{A}_{21}\mathbf{A}_{22}(\boldsymbol{\Lambda}_{123}+\boldsymbol{\Lambda}'_{123}),\\ \mathbf{Q}_{12}&=\mathbf{A}_{11}\mathbf{A}_{21}\boldsymbol{\Lambda}_{12}+\mathbf{A}_{12}\mathbf{A}_{22}\boldsymbol{\Lambda}_{33}+\mathbf{A}_{11}\mathbf{A}_{22}\boldsymbol{\Lambda}_{123}+\mathbf{A}_{12}\mathbf{A}_{21}\boldsymbol{\Lambda}'_{123},\\ \mathbf{Q}_{21}&=\mathbf{Q}'_{12}.\end{aligned} \tag{26}$$

Given the above normal distribution, the conditional distribution for $\boldsymbol{\zeta}_h$, a normal

$N(\boldsymbol{\zeta}_h \mid \boldsymbol{\theta}_h)$ can now be ascertained. Specifically,

$$\begin{aligned} \boldsymbol{\zeta}_h \mid \boldsymbol{\theta}_h &\sim N(E(\boldsymbol{\zeta}_h \mid \boldsymbol{\theta}_h), \mathrm{Var}(\boldsymbol{\zeta}_h \mid \boldsymbol{\theta}_h)), \\ E(\boldsymbol{\zeta}_h \mid \boldsymbol{\theta}_h) &= \mathbf{A}_{21}\mathbf{Z}_{h12}\boldsymbol{\alpha}_{12} + \mathbf{A}_{22}\mathbf{Z}_{h3}\boldsymbol{\alpha}_3 + \mathbf{Q}_{21}\mathbf{Q}_{11}^{-1}(\boldsymbol{\theta}_h - (\mathbf{A}_{11}\mathbf{Z}_{h12}\boldsymbol{\alpha}_{12} + \mathbf{A}_{12}\mathbf{Z}_{h3}\boldsymbol{\alpha}_3)), \\ \mathrm{Var}(\boldsymbol{\zeta}_h \mid \boldsymbol{\theta}_h) &= \mathbf{Q}_{22} - \mathbf{Q}_{21}\mathbf{Q}_{11}^{-1}\mathbf{Q}_{12}. \end{aligned} \tag{27}$$

6.2. *Full-Full-Choice*

In this information scenario, we have Full information in Category-1, Full information in Category-2 and Choice information in Category-3. In what follows, we show how to combine the available information across the three categories to make inferences about the household-specific parameters.

We begin with Category-3. The information set in this category contains only choices. The analysis described in the previous sub-section within Category-2 can be replicated in Category-3. Thus, we begin with the matrix representation for integrating the information across categories.

$$\begin{bmatrix} \boldsymbol{\beta}_{h1} \\ \boldsymbol{\beta}_{h2} \\ \boldsymbol{\mu}_{h3} \\ \boldsymbol{\delta}_{h3} \end{bmatrix} = \begin{bmatrix} \mathbf{I} & \mathbf{0} & \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \mathbf{I} & \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \mathbf{0} & \mathbf{I} & \mathbf{R}_3 \\ \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{I} \end{bmatrix} \begin{bmatrix} \boldsymbol{\beta}_{h1} \\ \boldsymbol{\beta}_{h2} \\ \boldsymbol{\gamma}_{h3} \\ \boldsymbol{\delta}_{h3} \end{bmatrix}. \tag{28}$$

This can be rewritten as:

$$\begin{bmatrix} \boldsymbol{\theta}_h \\ \boldsymbol{\delta}_{h3} \end{bmatrix} = \begin{bmatrix} \mathbf{A}_{11} & \mathbf{A}_{12} \\ \mathbf{A}_{21} & \mathbf{A}_{22} \end{bmatrix} \begin{bmatrix} \boldsymbol{\beta}_{h12} \\ \boldsymbol{\beta}_{h3} \end{bmatrix}, \tag{29}$$

where, $\boldsymbol{\theta}_h \equiv (\boldsymbol{\beta}'_{h1}, \boldsymbol{\beta}'_{h2}, \boldsymbol{\mu}'_{h3})'$, $\boldsymbol{\beta}_{h12} \equiv (\boldsymbol{\beta}'_{h1}, \boldsymbol{\beta}'_{h2})'$, $\boldsymbol{\beta}_{h3} \equiv (\boldsymbol{\gamma}'_{h3}, \boldsymbol{\delta}'_{h3})'$ and,

$$\begin{aligned} \mathbf{A}_{11} &= \begin{bmatrix} \mathbf{I} & \mathbf{0} \\ \mathbf{0} & \mathbf{I} \\ \mathbf{0} & \mathbf{0} \end{bmatrix}, \quad \mathbf{A}_{12} = \begin{bmatrix} \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \mathbf{0} \\ \mathbf{I} & \mathbf{R}_3 \end{bmatrix}, \\ \mathbf{A}_{21} &= [\mathbf{0} \quad \mathbf{0}], \quad \mathbf{A}_{22} = [\mathbf{0} \quad \mathbf{I}]. \end{aligned} \tag{30}$$

The reference group provides estimates of the population distribution's fixed effect $\boldsymbol{\alpha}$, and covariance $\boldsymbol{\Lambda}$.

Using the above defined submatrices of $\mathbf{A}$, we can rewrite the multivariate normal expression as

$$\begin{bmatrix} \boldsymbol{\theta}_h \\ \boldsymbol{\delta}_{h3} \end{bmatrix} \sim N\left(\begin{bmatrix} \mathbf{A}_{11} & \mathbf{A}_{12} \\ \mathbf{A}_{21} & \mathbf{A}_{22} \end{bmatrix}\begin{bmatrix} \mathbf{Z}_{h12}\boldsymbol{\alpha}_{12} \\ \mathbf{Z}_{h3}\boldsymbol{\alpha}_3 \end{bmatrix}, \begin{bmatrix} \mathbf{A}_{11} & \mathbf{A}_{12} \\ \mathbf{A}_{21} & \mathbf{A}_{22} \end{bmatrix}\begin{bmatrix} \boldsymbol{\Lambda}_{12} & \boldsymbol{\Lambda}_{123} \\ \boldsymbol{\Lambda}_{321} & \boldsymbol{\Lambda}_{33} \end{bmatrix}\begin{bmatrix} \mathbf{A}_{11} & \mathbf{A}_{21} \\ \mathbf{A}_{12} & \mathbf{A}_{22} \end{bmatrix}\right), \tag{31}$$

where, $\boldsymbol{\Lambda}_{12}$ refers to the cross-category covariance matrix between parameters in Category-1 and Category-2, $\boldsymbol{\Lambda}_{123}$ is the cross-category covariance matrix between parameters in Category-1, Category-2 and Category-3 and $\boldsymbol{\Lambda}_{33}$ represents the within category covariance matrix of the parameters in Category-3. Finally, $\mathbf{Z}_{h12}\boldsymbol{\alpha}_{12} \equiv (\mathbf{Z}_{h1}\boldsymbol{\alpha}_1', \mathbf{Z}_{h2}\boldsymbol{\alpha}_2')'$

For ease of illustration, we denote the covariance, $\mathbf{A}\boldsymbol{\Lambda}\mathbf{A}'$, as 2×2 matrix $\mathbf{Q}$, where,

$$\begin{aligned}
\mathbf{Q}_{11} &= \mathbf{A}_{11}^2\boldsymbol{\Lambda}_{12} + \mathbf{A}_{12}^2\boldsymbol{\Lambda}_{33} + \mathbf{A}_{11}\mathbf{A}_{12}(\boldsymbol{\Lambda}_{123} + \boldsymbol{\Lambda}_{123}'), \\
\mathbf{Q}_{22} &= \mathbf{A}_{21}^2\boldsymbol{\Lambda}_{12} + \mathbf{A}_{22}^2\boldsymbol{\Lambda}_{33} + \mathbf{A}_{21}\mathbf{A}_{22}(\boldsymbol{\Lambda}_{123} + \boldsymbol{\Lambda}_{123}'), \\
\mathbf{Q}_{12} &= \mathbf{A}_{11}\mathbf{A}_{21}\boldsymbol{\Lambda}_{12} + \mathbf{A}_{12}\mathbf{A}_{22}\boldsymbol{\Lambda}_{33} + \mathbf{A}_{11}\mathbf{A}_{22}\boldsymbol{\Lambda}_{123} + \mathbf{A}_{12}\mathbf{A}_{21}\boldsymbol{\Lambda}_{123}', \\
\mathbf{Q}_{21} &= \mathbf{Q}_{12}'.
\end{aligned} \tag{32}$$

Given the above normal distribution, the conditional distribution for $\boldsymbol{\delta}_{h3}$, a normal $N(\boldsymbol{\delta}_{h3} \mid \boldsymbol{\theta}_h)$ can now be ascertained. Specifically,

$$\begin{aligned}
\boldsymbol{\delta}_{h3} \mid \boldsymbol{\theta}_h &\sim N(E(\boldsymbol{\delta}_{h3} \mid \boldsymbol{\theta}_h), \mathrm{Var}(\boldsymbol{\delta}_{h3} \mid \boldsymbol{\theta}_h)), \\
E(\boldsymbol{\delta}_{h3} \mid \boldsymbol{\theta}_h) &= \mathbf{A}_{21}\mathbf{Z}_{h12}\boldsymbol{\alpha}_{12} + \mathbf{A}_{22}\mathbf{Z}_{h3}\boldsymbol{\alpha}_3 + \mathbf{Q}_{21}\mathbf{Q}_{11}^{-1}(\boldsymbol{\theta}_h - (\mathbf{A}_{11}\mathbf{Z}_{h12}\boldsymbol{\alpha}_{12} + \mathbf{A}_{12}\mathbf{Z}_{h3}\boldsymbol{\alpha}_3)), \\
\mathrm{Var}(\boldsymbol{\delta}_{h3} \mid \boldsymbol{\theta}_h) &= \mathbf{Q}_{22} - \mathbf{Q}_{21}\mathbf{Q}_{11}^{-1}\mathbf{Q}_{12}.
\end{aligned} \tag{33}$$

Thus, our model lends itself to an extension to three categories. However, as is evident, increasing the number of categories, increase both the complexity and the number of parameters to be estimated. As more categories are considered, it would become necessary to impose a structure on the parameter covariance matrix in order to keep the estimation procedure feasible. Examples of this structure include variance component analysis (Ainslie and Rossi, 1998) or a factor analytical structure (Deepak et al., 2002).

7. Multi-category data

In this paper we focused on prediction and parameter recovery within a category in the presence of missing information. Using these two criteria, we assessed the information gain while using data from another category. In our simulation study, we find that, in general, there is very little gain in prediction from a cross-category model when compared to a single-category analysis. This suggests that in any empirical application, where the likelihood provides a way to compare approaches, a cross-category analysis is unlikely to show any significant gains over a single-category analysis. In many situations, however, leveraging information across categories improves the recovery of underlying household-specific parameters.

While we focus on a specific facet of multi-category information and find its informational value questionable, it is clear that there are several substantive problems that cannot be addressed in the absence of multi-category information. For instance, Ainslie and Rossi (1998) focus on the presence of correlation in price-sensitivities across multiple categories. It would be impossible to address this substantive problem in the absence of multiple category information. Our focus is also on the correlation in cross-category price sensitivities but, rather than just its presence, we question whether that correlation can be leveraged for parameter recovery and choice prediction within a category. We find that a high correlation among parameters of two categories is a necessary but not sufficient condition for information transfer between two product categories.

Apart from the substantive problem of investigating the correlation in price sensitivities across categories, there are other issues that cannot be addressed without multi-category data. For instance, managers are interested in understanding how brand equity can be leveraged across categories. Several studies have proposed that a common brand name helps in transferring quality across product categories (Aaker and Keller, 1990; Wernerfelt, 1988). This effect, called umbrella branding, has subsequently been empirically studied. Both Erdem (1998) and Manchanda et al. (2000) found that common brand names across categories did indeed help in transferring quality and that there was a strong correlation in the own price sensitivity across categories. Manchanda et al. (2000) also found that the cross-category effects were stronger for common brands. In a similar context, Erdem and Sun (2002) found evidence for advertising and promotion spill over effects for common brands across categories. All the empirical studies described above used multi-category data and it would be impossible to address these substantive issues in absence of such multi-category information.

Another stream multi-category research has focused on understanding the nature of relationships in cross-category purchasing. An interest in uncovering complementary and substitute relations between products has its roots in both marketing (Bass et al., 1969) and economics (Stone, 1954). The identification of substitute and complementary products can help in understanding how shopping baskets are put together. This understanding, besides being of obvious interest from an academic perspective, can guide managers to forecast the cross-category effects of any marketing action. In this vein, Manchanda et al. (1999) used multi-category grocery data to separate the effects of product complementarity from that of co-occurrence or coincidence on the formation of a shopping basket. As with the substantive issue of umbrella branding, the issues of complementarity and substitution cannot be suitably addressed without multi-category data. Thus, we are certainly not making a claim that multi-category information is questionable under all circumstances.

In sum, the value of cross-category data depends on the objective of the researcher. Our focus is well suited for providing guidelines to managers for implementing their micro marketing programs. We have shown conditions where a cross-category analysis can lead to better recovery of household-specific parameters as compared to a single category analysis. Any decision problem that relies on household-specific parameter estimates to derive optimal managerial actions could directly benefit from an improved recovery of parameters. For instance, in optimal couponing decisions (Rossi et al., 1996), improved estimation of household-specific parameters can alter the distribution of household-specific optimal coupons in a target category. However, we also showed that there are specific conditions that are conducive for this better parameter recovery. Thus, managers have to carefully consider the cost–benefit tradeoffs associated with collecting and utilizing multi-category information.

8. Conclusion

As companies collect increasing amounts of information about their customers, there is a strong need to understand how this information can be leveraged to predict consumers' behavior in multiple categories. This could help companies in many situations such as cross-selling. Building on the work of Rossi et al. (1996) who show the value of purchase information within a single product category, we show how information can be leveraged across multiple categories when a firm has limited or no information about a target customer in one or more categories.

To understand the boundary conditions of information transfer across categories, we conduct an extensive simulation that shows many interesting and surprising findings. For example, we find that in most situations a single-category model is likely to give rise to data likelihood as good as a cross-category model. This makes the work of Rossi et al. (1996) even more powerful than previously recognized. However, our simulation also shows the conditions when additional information is likely to be beneficial. Specifically, we find that there are a few conditions wherein there is a decrease in the ESS when a cross-category model is considered as compared to a single-category analysis. In those conditions, the distribution of household-specific optimal coupons in a target category would be different if cross-category information is considered as opposed to household information from only the target category.

The guidelines developed from this simulation help us to predict the outcome for empirical applications. We confirm this in our application involving two related grocery products—Breakfast Foods and Table Syrup. In spite of a reasonable correlation in the price parameter across these two categories, a cross-category analysis provides no significant benefit over a single-category analysis, consistent with our simulation guidelines.

There are several limitations of the present work and these provide avenues for future research.[5] In this paper, we show that the information content of cross-category data is questionable. We, however, empirically show this result with only two categories. It is possible that our results might change if more categories are considered. This would be the case as with more categories there is greater information about household-level purchase behavior and this might aid in leveraging any cross-category correlation. In our model, we utilize an unstructured covariance matrix between the cross-category parameters. On the one hand, an unstructured covariance matrix is flexible but on the other, as more categories are considered, it suffers from the curse of dimensionality. It would be interesting to explore whether putting a structure on the covariance matrix, and thereby relatively easily accommodating more categories, helps in better parameter recovery and/or prediction.

Appendix A. Full conditional distributions for the cross-category model

Markov chain Monte Carlo procedures are used for numerical Bayesian inference for the reference group data. MCMC involves iterative sampling from the full conditional distributions. The multiple iterations of the MCMC procedure then generate a Markov chain that converges in distribution to the joint posterior under fairly general conditions. Thus, the essential idea is to construct a chain whose stationary distribution is the required posterior distribution. We extend to a cross-category model the MCMC procedure for the multinomial probit model (Albert and Chib, 1993; McCulloch and Rossi, 1994) to generate the sample of draws. The following set of full conditionals and priors for $\boldsymbol{\alpha}, \boldsymbol{\Lambda}, \Sigma_1, \Sigma_2, \boldsymbol{\beta}_{h1}, \boldsymbol{\beta}_{h2}$ and $\mathbf{u}$ are used:

a. The full conditional for $\boldsymbol{\alpha}$ is multivariate normal $N(\bar{\boldsymbol{\alpha}}, \mathbf{V}_\alpha)$, where $\mathbf{V}_\alpha^{-1} = \mathbf{C}^{-1} + \sum_{h=1}^{H} \mathbf{Z}_h' \boldsymbol{\Lambda}^{-1} \mathbf{Z}_h$ and $\bar{\boldsymbol{\alpha}} = \mathbf{V}_\alpha[\mathbf{C}^{-1}\boldsymbol{\eta} + \sum_{h=1}^{H} \mathbf{Z}_h' \boldsymbol{\Lambda}^{-1} \boldsymbol{\beta}_h]$. The prior mean $\boldsymbol{\eta}$ was fixed to be $\mathbf{0}$ and $\mathbf{C}$ was set as a diagonal matrix diag(1000).
b. The full conditional for the precision matrix $\boldsymbol{\Lambda}^{-1}$ of the population distribution is a Wishart distribution. The full conditional can be written as $W(\rho_\lambda, \mathbf{V}_\lambda)$, where $\mathbf{V}_\lambda = (\sum_{h=1}^{H} (\boldsymbol{\beta}_h - \mathbf{Z}_h\boldsymbol{\alpha})(\boldsymbol{\beta}_h - \mathbf{Z}_h\boldsymbol{\alpha})' + \rho\boldsymbol{\Gamma})^{-1}$ and $\rho_\lambda = \rho + H$. H is the number of households, ρ which is the df for the Wishart prior is set to $N_p + 1$, where N_p is the dimension of $\boldsymbol{\beta}_h$ and $\boldsymbol{\Gamma}$ the scale matrix for the Wishart prior is fixed at identity.
c. Draws for $\boldsymbol{\beta}_h$ are obtained by obtaining conditional draws for $\boldsymbol{\beta}_{h1}$ and $\boldsymbol{\beta}_{h2}$. We consider the full conditional of $\boldsymbol{\beta}_{h1}$ here and the one for $\boldsymbol{\beta}_{h2}$ is analogous. The prior for $\boldsymbol{\beta}_h$ is $N(\mathbf{Z}_h\boldsymbol{\alpha}, \boldsymbol{\Lambda})$. We can derive the conditional prior distribution of $\boldsymbol{\beta}_{h1}$

5 We thank the reviewers for bringing a few of these to our attention.

given $\boldsymbol{\beta}_{h2}$. Specifically,

$$\begin{aligned}\boldsymbol{\beta}_{h1} \mid \boldsymbol{\beta}_{h2} &\sim N(E(\boldsymbol{\beta}_{h1} \mid \boldsymbol{\beta}_{h2}), \mathrm{Var}(\boldsymbol{\beta}_{h1} \mid \boldsymbol{\beta}_{h2})),\\ E(\boldsymbol{\beta}_{h1} \mid \boldsymbol{\beta}_{h2}) &= \mathbf{Z}_{h1}\boldsymbol{\alpha}_1 + \boldsymbol{\Lambda}_{12}\boldsymbol{\Lambda}_{22}^{-1}(\boldsymbol{\beta}_{h2} - \mathbf{Z}_{h2}\boldsymbol{\alpha}_2),\\ \mathrm{Var}(\boldsymbol{\beta}_{h1} \mid \boldsymbol{\beta}_{h2}) &= \boldsymbol{\Lambda}_{11} - \boldsymbol{\Lambda}_{12}\boldsymbol{\Lambda}_{22}^{-1}\boldsymbol{\Lambda}_{21},\end{aligned} \tag{1}$$

where, $\boldsymbol{\Lambda}_{11}$ is the covariance among the parameters of Category-1, $\boldsymbol{\Lambda}_{22}$ is the covariance among the parameters in Category-2 and $\boldsymbol{\Lambda}_{12}$ refers to the covariance between the parameters of Category-1 and Category-2.

For ease of exposition, we can write the conditional normal prior as $p(\boldsymbol{\beta}_{h1} \mid \boldsymbol{\beta}_{h2}) \sim N(\boldsymbol{\alpha}_{1|2}, \boldsymbol{\Lambda}_{1|2})$. With this conditional prior, the full conditional of $\boldsymbol{\beta}_{h1}$ is normal, $N(\bar{\boldsymbol{\beta}}_{h1}, \mathbf{V}_{\beta_{h1}})$ with $\mathbf{V}_{\beta_{h1}}^{-1} = \boldsymbol{\Lambda}_{1|2}^{-1} + \sum_{t=1}^{n_{h1}} \mathbf{X}'_{ht1}\Sigma_1^{-1}\mathbf{X}_{ht1}$ and $\bar{\boldsymbol{\beta}}_{h1}$ is $\mathbf{V}_{\boldsymbol{\beta}_{h1}}[\boldsymbol{\Lambda}_{1|2}^{-1}\boldsymbol{\alpha}_{h1\,|\,2} + \sum_{t=1}^{n_{h1}} \mathbf{X}'_{ht1}\Sigma_1^{-1}\mathbf{u}_{ht1}]$.

d. The full conditional distribution for $\boldsymbol{\Sigma}_1$ and $\boldsymbol{\Sigma}_2$ can be considered separately given their assumed independence. Here we consider the full conditional of $\boldsymbol{\Sigma}_1$. The full conditional of $\boldsymbol{\Sigma}_2$ is analogous. $\boldsymbol{\Sigma}_1$ is diagonal with the first element set to 1 (for scaling purposes). As the utilities within Category-1 are considered independent given $\boldsymbol{\beta}_{h1}$, each element of $\boldsymbol{\Sigma}_1$ can be considered separately. Considering the kth element of $\boldsymbol{\Sigma}_1$, we set an inverse gamma prior $IG(a,b)$. Thus, $p(\sigma_{1k}) \sim IG(a,b)$. With this prior, the posterior is $IG(N_1/2 + a, [1/2\sum_{h=1}^{H}\sum_{t=1}^{n_{1h}}(u_{htk1} - \mathbf{X}'_{htk1}\boldsymbol{\beta}_{h1})^2 + b^{-1}]^{-1})$ where N_1 is the number of observations within Category-1. We fixed a at 3 and b at 0.5 for the priors.
e. The utilities $\mathbf{u}_{ht1}$ and $\mathbf{u}_{ht2}$ for any given observation in each category can be drawn in a data augmentation step as detailed in Rossi et al. (1996). This involves drawing each brand utility from a truncated conditional normal distribution where the truncation points depend upon whether the utility is for the chosen brand or not.

Appendix B. Information scenarios

In this appendix we show how to make inferences for the household-specific parameters, $\boldsymbol{\beta}_h$, for the target group under the different Scenarios.

B.1. Scenario-1 (Base-Base)

In this Scenario, no specific information about any target household h is available in both categories. Inferences about the response coefficients $\boldsymbol{\beta}_h$ can be made by utilizing the information about the reference group in both categories. Specifically, the reference group gives information about the population parameters, $\boldsymbol{\alpha}$ and $\boldsymbol{\Lambda}$.

For any particular target household h, we can write,

$$p(\boldsymbol{\beta}_h \mid D) = \int p(\boldsymbol{\beta}_h \mid \mathbf{Z}_h, \boldsymbol{\alpha}, \boldsymbol{\Lambda}) f(\mathbf{Z}_h) f(\boldsymbol{\alpha}, \boldsymbol{\Lambda} \mid D)\, \mathrm{d}\mathbf{Z}_h\, \mathrm{d}\boldsymbol{\alpha}\, \mathrm{d}\boldsymbol{\Lambda}, \tag{1}$$

where $p(\boldsymbol{\beta}_h \mid D)$ is the predictive distribution conditional on the data D of the reference group, $\mathbf{Z}_h$ represents the block-diagonal matrix of demographics influencing $\boldsymbol{\beta}_h$, $f(\mathbf{Z}_h)$ represents a distribution of the demographics and $f(\boldsymbol{\alpha}, \boldsymbol{\Lambda} \mid D)$ is the joint posterior distribution of population parameters conditional on D.

As the demographics of a particular household, i.e., $\mathbf{Z}_h$, are unknown, these have to be averaged out. This can be done by assuming a suitable demographics distribution, $f(\mathbf{Z}_h)$, such as the empirical distribution in the data, and then sampling from that distribution. The integration over the population parameters, $\boldsymbol{\alpha}$ and $\boldsymbol{\Lambda}$, can be carried out by using samples obtained from the joint posterior distribution as a consequence of estimating the cross-category model on the reference group. This procedure gives a predictive distribution for parameters in both Category-1 and Category-2. Rossi et al. (1996) employ a similar inference procedure. They, however, use information of the reference group from Category-2 to infer the parameters of the target group in that category. We differ by using information of the reference group in both categories. This helps to leverage any information that is present in the cross-category covariance between the parameters.

B.2. Scenario-2 (Demo-Demo)

This Scenario builds on the previous one by letting retailers have knowledge about demographic attributes of their customer base. In the context of making inferences, the only difference from Scenario-1 is that there is no need to average out the distribution of demographics. Thus,

$$p(\boldsymbol{\beta}_h \mid \mathbf{Z}_h, D) = \int p(\boldsymbol{\beta}_h \mid \boldsymbol{\alpha}, \boldsymbol{\Lambda}) p(\boldsymbol{\alpha}, \boldsymbol{\Lambda} \mid D)\, \mathrm{d}\boldsymbol{\alpha}\, \mathrm{d}\boldsymbol{\Lambda}, \tag{2}$$

where $p(\boldsymbol{\beta}_h \mid \mathbf{Z}_h, D)$is the predictive distribution conditional on the data D of the reference group and the known demographics matrix, $\mathbf{Z}_h$ of the target household and $f(\boldsymbol{\alpha}, \boldsymbol{\Lambda} \mid D)$represents the joint posterior distribution of population parameters conditional on the reference group data D.

B.3. Scenario-3 (Demo-Choice)

In this Scenario, the demographics of the target household and choices in one category (without loss of generalization Category-1) are known. Scenario-2 provides the methodology for estimating the target household parameters when only demographics of the target households are known. The new element in this Scenario is that choice information is available within Category-1 and no purchase information is available for Category-2. In what follows, we adapt the methodology outlined in Rossi et al. (1996) when only choice information is available within a category to our cross-category context.

For the target households, we can only run an intercept-level model within Category-1. This yields

$$\mathbf{u}_{ht1} = \begin{bmatrix} 0 \\ \boldsymbol{\mu}_{h1} \end{bmatrix} + \mathbf{e}_{ht1}. \tag{3}$$

We have to map the posterior distribution of $\boldsymbol{\mu}_{h1}$ into a distribution on $\boldsymbol{\beta}_{h1}$ and need to simultaneously infer $\boldsymbol{\beta}_{h2}$ based on the demographics for household h and the population parameters from the reference group. The complete model for households within Category-1 can be written as

$$\mathbf{u}_{ht1} = \mathbf{X}_{ht1}\boldsymbol{\beta}_{h1} + \mathbf{e}_{ht1} = \begin{bmatrix} 0 \\ \boldsymbol{\gamma}_{h1} \end{bmatrix} + \mathbf{X}_{cht1}\boldsymbol{\delta}_{h1} + \mathbf{e}_{ht1}, \tag{4}$$

where, $\mathbf{X}_{cht1}$ is a matrix containing causal variables, which is not available for the target households. Therefore, to map $\boldsymbol{\mu}_{h1}$ into $\boldsymbol{\beta}_{h1}$, following Rossi et al. (1996), we can take iterated expectations and equate these expectations with the intercepts obtained in equation (3) above. Formally, $E[u] = E_x[E[u \mid \mathbf{X}]] = E_x[\mathbf{X}\boldsymbol{\beta}]$ and thus,

$$E[u] = \begin{bmatrix} \overline{\mathbf{x}'_1}\boldsymbol{\delta}_{h1} \\ \gamma_{21} + \overline{\mathbf{x}'_2}\boldsymbol{\delta}_{h1} \\ \vdots \\ \gamma_{p1} + \overline{\mathbf{x}'_p}\boldsymbol{\delta}_{h1} \end{bmatrix}. \tag{5}$$

In the above equation, $\overline{\mathbf{x}'_j}$ is the vector of means of the causal variables for the jth alternative. It is through these means that the prior information on the distribution of causal variables is incorporated. After making adjustments for differences in identification, we can write in matrix form,

$$\boldsymbol{\gamma}_{h1} + \mathbf{R}_1\boldsymbol{\delta}_{h1} = \boldsymbol{\mu}_{h1}, \tag{6}$$

where,

$$\mathbf{R}_1 = \begin{bmatrix} \overline{x'_2} - \overline{x'_1} \\ \overline{x'_3} - \overline{x'_1} \\ \vdots \\ \overline{x'_n} - \overline{x'_1} \end{bmatrix}.$$

This analysis needs to be integrated with the reference group information in both categories. Conceptually, these sources of information can be brought together as a set of equations,

$$\begin{bmatrix} \boldsymbol{\mu}_{h1} \\ \boldsymbol{\delta}_{h1} \\ \boldsymbol{\beta}_{h2} \end{bmatrix} = \begin{bmatrix} \mathbf{I} & \mathbf{R}_1 & \mathbf{0} \\ \mathbf{0} & \mathbf{I} & \mathbf{0} \\ \mathbf{0} & \mathbf{0} & \mathbf{I} \end{bmatrix} \begin{bmatrix} \boldsymbol{\gamma}_{h1} \\ \boldsymbol{\delta}_{h1} \\ \boldsymbol{\beta}_{h2} \end{bmatrix}. \tag{7}$$

Rewriting $(\boldsymbol{\delta}'_{h1}, \boldsymbol{\beta}'_{h2})$as $\boldsymbol{\theta}'_h$, we have

$$\begin{bmatrix} \boldsymbol{\mu}_{h1} \\ \boldsymbol{\theta}_h \end{bmatrix} = [\mathbf{A}] \begin{bmatrix} \boldsymbol{\gamma}_{h1} \\ \boldsymbol{\delta}_{h1} \\ \boldsymbol{\beta}_{h2} \end{bmatrix}, \tag{8}$$

where,

$$\mathbf{A} = \begin{bmatrix} \mathbf{I} & \mathbf{R}_1 & \mathbf{0} \\ \mathbf{0} & \mathbf{I} & \mathbf{0} \\ \mathbf{0} & \mathbf{0} & \mathbf{I} \end{bmatrix}.$$

Further, we can write $(\boldsymbol{\gamma}'_{h1}, \boldsymbol{\delta}'_{h1}) \equiv \boldsymbol{\beta}'_{h1}$ to obtain

$$\begin{bmatrix} \boldsymbol{\mu}_{h1} \\ \boldsymbol{\theta}_h \end{bmatrix} = [\mathbf{A}] \begin{bmatrix} \boldsymbol{\beta}_{h1} \\ \boldsymbol{\beta}_{h2} \end{bmatrix}. \tag{9}$$

The reference group provides estimates of the population distribution's fixed effect $\boldsymbol{\alpha}$, and covariance $\boldsymbol{\Lambda}$. These estimates allow us to write the following

$$\begin{bmatrix} \boldsymbol{\mu}_{h1} \\ \boldsymbol{\theta}_h \end{bmatrix} \sim N(\mathbf{AZ}\boldsymbol{\alpha}, \mathbf{A}\boldsymbol{\Lambda}\mathbf{A}'). \tag{10}$$

For ease of explication, let,

$$\mathbf{A} = \begin{bmatrix} \mathbf{A}_{11} & \mathbf{A}_{12} \\ \mathbf{A}_{21} & \mathbf{A}_{22} \end{bmatrix}, \tag{11}$$

where,

$$\mathbf{A}_{11} = [\mathbf{I} \quad \mathbf{R}_1], \quad \mathbf{A}_{12} = [\mathbf{0}], \quad \mathbf{A}_{21} = \begin{bmatrix} \mathbf{0} & \mathbf{I} \\ \mathbf{0} & \mathbf{0} \end{bmatrix}, \quad \mathbf{A}_{22} = \begin{bmatrix} \mathbf{0} \\ \mathbf{I} \end{bmatrix}.$$

Using the above defined submatrices of **A**, we can rewrite the multivariate normal expression as

$$\begin{bmatrix} \boldsymbol{\mu}_{h1} \\ \boldsymbol{\theta}_h \end{bmatrix} \sim N\left(\begin{bmatrix} \mathbf{A}_{11} & \mathbf{A}_{12} \\ \mathbf{A}_{21} & \mathbf{A}_{22} \end{bmatrix} \begin{bmatrix} \mathbf{Z}_{h1}\boldsymbol{\alpha}_1 \\ \mathbf{Z}_{h2}\boldsymbol{\alpha}_2 \end{bmatrix}, \begin{bmatrix} \mathbf{A}_{11} & \mathbf{A}_{12} \\ \mathbf{A}_{21} & \mathbf{A}_{22} \end{bmatrix} \begin{bmatrix} \boldsymbol{\Lambda}_{11} & \boldsymbol{\Lambda}_{12} \\ \boldsymbol{\Lambda}_{21} & \boldsymbol{\Lambda}_{22} \end{bmatrix} \begin{bmatrix} \mathbf{A}_{11} & \mathbf{A}_{21} \\ \mathbf{A}_{12} & \mathbf{A}_{22} \end{bmatrix} \right). \tag{12}$$

For ease of illustration, we denote the covariance, $\mathbf{A\Lambda A}'$, as 2×2 matrix $\mathbf{Q}$, where,

$$\begin{aligned} \mathbf{Q}_{11} &= \mathbf{A}_{11}^2\boldsymbol{\Lambda}_{11} + \mathbf{A}_{12}^2\boldsymbol{\Lambda}_{22} + \mathbf{A}_{11}\mathbf{A}_{12}(\boldsymbol{\Lambda}_{12} + \boldsymbol{\Lambda}_{21}), \\ \mathbf{Q}_{22} &= \mathbf{A}_{21}^2\boldsymbol{\Lambda}_{11} + \mathbf{A}_{22}^2\boldsymbol{\Lambda}_{22} + \mathbf{A}_{21}\mathbf{A}_{22}(\boldsymbol{\Lambda}_{12} + \boldsymbol{\Lambda}_{21}), \\ \mathbf{Q}_{12} &= \mathbf{A}_{11}\mathbf{A}_{21}\boldsymbol{\Lambda}_{11} + \mathbf{A}_{12}\mathbf{A}_{22}\boldsymbol{\Lambda}_{22} + \mathbf{A}_{11}\mathbf{A}_{22}\boldsymbol{\Lambda}_{12} + \mathbf{A}_{12}\mathbf{A}_{21}\boldsymbol{\Lambda}_{21}, \\ \mathbf{Q}_{21} &= \mathbf{Q}_{12}'. \end{aligned} \tag{13}$$

Given the above normal distribution, the conditional distribution for $\boldsymbol{\theta}_h$, a normal $N(\boldsymbol{\theta}_h \mid \boldsymbol{\mu}_{h1})$can now be ascertained. Specifically,

$$\begin{aligned} \boldsymbol{\theta}_h \mid \boldsymbol{\mu}_{h1} &\sim N(E(\boldsymbol{\theta}_h \mid \boldsymbol{\mu}_{h1}), \mathrm{Var}(\boldsymbol{\theta}_h \mid \boldsymbol{\mu}_{h1})), \\ E(\boldsymbol{\theta}_h \mid \boldsymbol{\mu}_{h1}) &= \mathbf{A}_{21}\mathbf{Z}_{h1}\boldsymbol{\alpha}_1 + \mathbf{A}_{22}\mathbf{Z}_{h2}\boldsymbol{\alpha}_2 + \mathbf{Q}_{21}\mathbf{Q}_{11}^{-1}(\boldsymbol{\mu}_{h1} - (\mathbf{A}_{11}\mathbf{Z}_{h1}\boldsymbol{\alpha}_1 + \mathbf{A}_{12}\mathbf{Z}_{h2}\boldsymbol{\alpha}_2)), \\ \mathrm{Var}(\boldsymbol{\theta}_h \mid \boldsymbol{\mu}_{h1}) &= \mathbf{Q}_{22} - \mathbf{Q}_{21}\mathbf{Q}_{11}^{-1}\mathbf{Q}_{12}. \end{aligned} \tag{14}$$

B.4. Scenario-4 (Base-Full)

This Scenario is discussed in detail in the text.

B.5. Scenario-5 (Choice-Choice)

This Scenario is discussed in detail in the text.

B.6. Scenario-6 (Choice-Full)

In this Scenario, we have full information in one category (without loss of generality, Category-1) and only choice in the other (Category-2). Thus for the target group, we can estimate a cross-category probit model that yields the response parameters $\boldsymbol{\beta}_{h1}$ in Category-1 and the intercepts $\boldsymbol{\mu}_{h2}$ in Category-2. Using these estimates and the information from the reference group, the task is to obtain estimates for response coefficients for the causal variables in Category-2, i.e., $\boldsymbol{\delta}_{h2}$. We can write the relationship between these three sets of parameters in matrix form as follows:

$$\begin{bmatrix} \boldsymbol{\beta}_{h1} \\ \boldsymbol{\mu}_{h2} \\ \boldsymbol{\delta}_{h2} \end{bmatrix} = \begin{bmatrix} \mathbf{I} & \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \mathbf{I} & \mathbf{R}_2 \\ \mathbf{0} & \mathbf{0} & \mathbf{I} \end{bmatrix} \begin{bmatrix} \boldsymbol{\beta}_{h1} \\ \boldsymbol{\gamma}_{h2} \\ \boldsymbol{\delta}_{h2} \end{bmatrix}, \tag{15}$$

where $\mathbf{R}_2$ is analogous to that defined previously. We denote,

$$\mathbf{A} = \begin{bmatrix} \mathbf{I} & \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \mathbf{I} & \mathbf{R}_2 \\ \mathbf{0} & \mathbf{0} & \mathbf{I} \end{bmatrix}. \tag{16}$$

For ease of explication, let,

$$\mathbf{A} = \begin{bmatrix} \mathbf{A}_{11} & \mathbf{A}_{12} \\ \mathbf{A}_{21} & \mathbf{A}_{22} \end{bmatrix}, \tag{17}$$

where,

$$\mathbf{A}_{11} = \begin{bmatrix} \mathbf{I} \\ \mathbf{0} \end{bmatrix}, \quad \mathbf{A}_{12} = \begin{bmatrix} \mathbf{0} & \mathbf{0} \\ \mathbf{I} & \mathbf{R}_2 \end{bmatrix}, \quad \mathbf{A}_{21} = [\mathbf{0}], \quad \text{and } \mathbf{A}_{22} = [\mathbf{0} \quad \mathbf{I}].$$

The reference group provides estimates of the population distribution's fixed effect $\boldsymbol{\alpha}$, and covariance $\mathbf{\Lambda}$. These estimates allow us to write the following

$$\begin{bmatrix} \boldsymbol{\theta}_h \\ \boldsymbol{\delta}_{h2} \end{bmatrix} \sim N(\mathbf{A}\mathbf{Z}_h\boldsymbol{\alpha}, \mathbf{A}\mathbf{\Lambda}\mathbf{A}'), \tag{18}$$

where, $\boldsymbol{\theta}_h \equiv (\boldsymbol{\beta}'_{h1}, \boldsymbol{\mu}'_{h2})'$.

Using the above defined submatrices of $\mathbf{A}$, we can rewrite the expression as

$$\begin{bmatrix} \boldsymbol{\theta}_h \\ \boldsymbol{\delta}_{h2} \end{bmatrix} \sim N\left(\begin{bmatrix} \mathbf{A}_{11} & \mathbf{A}_{12} \\ \mathbf{A}_{21} & \mathbf{A}_{22} \end{bmatrix} \begin{bmatrix} \mathbf{Z}_{h1}\boldsymbol{\alpha}_1 \\ \mathbf{Z}_{h2}\boldsymbol{\alpha}_2 \end{bmatrix}, \begin{bmatrix} \mathbf{A}_{11} & \mathbf{A}_{12} \\ \mathbf{A}_{21} & \mathbf{A}_{22} \end{bmatrix} \begin{bmatrix} \mathbf{\Lambda}_{11} & \mathbf{\Lambda}_{12} \\ \mathbf{\Lambda}_{21} & \mathbf{\Lambda}_{22} \end{bmatrix} \begin{bmatrix} \mathbf{A}_{11} & \mathbf{A}_{21} \\ \mathbf{A}_{12} & \mathbf{A}_{22} \end{bmatrix} \right). \tag{19}$$

For ease of illustration, we denote the covariance, $\mathbf{A\Lambda A}'$, as 2×2 matrix $\mathbf{Q}$, with components $\mathbf{Q}_{11}, \mathbf{Q}_{12}, \mathbf{Q}_{21}$ and $\mathbf{Q}_{22}$ defined in equation (13) in the appendix. As in Scenario-3, the conditional distribution for $\boldsymbol{\delta}_{2h}$, is a normal $N(\boldsymbol{\delta}_{h2} \mid \boldsymbol{\theta}_h)$where the mean and variance of this distribution can be obtained from equation (14) of the appendix.

B.7. Scenario-7 (Full-Full)

In this information set, full information is available for every target household h in both categories. Hence, there is no difference between the reference group and target group in terms of the level of available information. Thus, within each category, the reference group and the target group can be combined and our cross-category model can be estimated on these combined groups. As a by product of the estimation procedure, we can obtain the inferences for the household-level parameters.

References

Aaker, David A. and Kevin K. Keller. (1990). "Consumer Evaluations of Brand Extensions", Journal of Marketing 54, 27–41.

Albert, James H. and Siddhartha Chib. (1993). "Bayesian Analysis of Binary and Polychotomous Response Data", Journal of American Statistical Association 88, 669–679.

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

Bass, Frank M., Edgar A. Pessemier, and Douglas J. Tigert. (1969). "Complementary and Substitute Patterns of Purchasing and Use", Journal of Advertising Research 9, 19–27.

Chib, Siddhartha, Edward Greenberg, and Yuxin Chen. (1998). "MCMC Methods for Fitting and Comparing Multinomial Response Models", Working Paper, Washington University, St. Louis.

Deepak, Sri Devi, Asim Ansari, and Sunil Gupta. (2002). "Investigating Consumer Price Sensitivities Across Categories", Working Paper, Columbia University.

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

Erdem, Tülin and Baohung Sun. (2002). ''An Empirical Investigation of the Spillover Effects of Advertising and Sales Promotions in Umbrella Branding'', Journal of Marketing Research 39, 408–420.
Erdem, Tülin, Glenn Mayhew, and Baohong Sun. (2001). ''Understanding the Reference Price Sensitive Shopper: A Within and Cross-Category Analysis'', Journal of Marketing Research 38, 445–457.
Gelfand, Alan E. and Adrian F.M. Smith. (1990). ''Sampling Based Approaches to Calculating Marginal Densities'', Journal of American Statistical Association 85, 972–985.
Geweke, J. (1991). ''Efficient Simulations for the Multivariate Normal and Student-t Distributions Subject to Linear Constraints'', Computer Science and Statistics: Proceedings of the Twenty-Third Symposium on the Interface, 571–578.
Hajivassiliou, V.A. (1990). ''Smooth Simulation Estimation of Panel LDV Models'', manuscript.
Kamakura, Wagner, Michel Wedel, Fernando de Rosa, and J.A. Mazzon. (2003). ''Cross-Selling Through Database Marketing: A Mixed Data Factor Analyzer for Data Augmentation and Prediction'', International Journal of Research in Marketing 20, 45–65.
Kasulis, J.J., R.F. Lusch, and E.F. Stafford. (1979). ''Consumer Acquisition Patterns for Durable Goods'', Journal of Consumer Research 6, 47–57.
Keane, Michael P. (1994). ''A Computationally Practical Simulation Estimator for Panel Data'', Econometrica 62, 95–116.
Manchanda, Puneet, Asim Ansari, and Sunil Gupta. (2000). ''Multi-Category Brand Effects'', Working Paper, Columbia University.
Manchanda, Puneet, Asim Ansari, and Sunil Gupta. (1999). ''The 'Shopping Basket': A Model for Multicategory Purchase Incidence Model'', Marketing Science 18, 95–114.
McCulloch, R.E. and Peter Rossi. (1994). ''An Exact Likelihood Analysis of the Multinomial Probit Model'', Journal of Econometrics Sept.–Oct., 207–40.
Rossi, Peter E., Robert E. McCulloch, and Greg M. Allenby. (1996). ''The Value of Purchase History Data in Target Marketing'', Marketing Science 321–340.
Stone, J.R.N. (1954). ''Linear Expenditure Demand Systems and Demand Analysis: An Application to the Pattern of British Demand'', Economic Journal 64, 511–527.
Wedel, Michel and Kamakura, Wagner. (1997). ''Statistical Data Fusion for Cross-tabulation'', Journal of Marketing Research 34, 485–498.
Wernerfelt, Birger. (1988). ''Umbrella Branding as a Signal of New Product Quality: An Example of Signaling by Posting a Bond'', RAND Journal of Economics 19, 458–466.