Modeling Interdependent Consumer Preferences

Sha Yang
Stern School of Business
New York University
44 West Fourth Street, Suite 9-77
New York, NY 10012
Tel: 212-998-0519
shayang@stern.nyu.edu

Greg M. Allenby
Fisher College of Business
The Ohio State University
2100 Neil Avenue
Columbus, OH 43210-1144
Tel: 614-292-9452
allenby.1@osu.edu

November, 2001
July, 2002
September, 2002

We thank the Editor and reviewers for their constructive comments and suggestions. We also thank seminar participants at University of California – Berkeley, New York University, Cornell University, Hong Kong University of Science and Technology, National University of Singapore, University of Connecticut, Case Western Reserve University, and the University of Wisconsin at Madison for their helpful comments.
Modeling Interdependent Consumer Preferences

Abstract

An individual's preference for an offering can be influenced by the preferences of others in many ways, ranging from the influence of social identification and inclusion, to the benefits of network externalities. In this paper, we introduce a Bayesian autoregressive discrete choice model to study the preference interdependence among individual consumers. The autoregressive specification can reflect patterns of heterogeneity where influence propagates within and across networks. These patterns cannot be modeled with standard random-effect specifications, and can be difficult to capture with covariates in a linear model. Our model of interdependent preferences is illustrated with data on automobile purchases, where preferences for Japanese made cars are shown to be related to geographic and demographically defined networks.
Modeling Interdependent Consumer Preferences

1. Introduction

Preferences and choice behavior are influenced by a consumer's own tastes and also the tastes of others. People who identify with a particular group often adopt the preferences of the group, resulting in choices that are interdependent. Examples include the preference for particular brands (e.g. Abercrombie and Fitch) and even entire product categories (e.g. minivans). Interdependence may be driven by social concerns, by endorsements from respected individuals that increase a brand's credibility, or by learning the preference of others who may have information not available to the decision maker. Moreover, since people engage in multiple activities with their families, co-workers, neighbors and friends, interdependent preferences can propagate across and through multiple networks.

Quantitative models of consumer purchase behavior often do not recognize that preferences and choices are interdependent. Economic models of choice typically assume that an individual's latent utility is a function of brand and attribute preferences, not the preferences of others. Preferences are assumed to vary across consumers in a manner described either by exogenous covariates, such as demographics (e.g., household income), or by independent draws from a mixing distribution in random-effects models (see Allenby and Rossi 1999, Kamakura and Russell 1989). However, if preferences in a market are interdependently determined, their pattern will not be well represented by a simple linear model of exogenous covariates. Failure to include high-order interaction terms in the model to reflect interdependent influences will result in correlated structure of unobserved heterogeneity, where the draws from a random-effects mixing distribution are dependent, not independent.
In this paper, we introduce a Bayesian model of interdependent preferences in a consumer choice context. We employ a parsimonious autoregressive structure that captures the endogenous relationship between an individual’s preference and the preference of others in the same network. Our model allows for a complex network associated with multiple explanatory variables. In addition, our model allows us to incorporate explanatory variables in both an exogenous and endogenous (i.e., interdependent) manner. A Markov chain Monte Carlo method is employed to estimate the model parameters.

The remainder of the paper is organized as follows. Section 2 reviews the literature in economics and marketing on interdependent preference, providing rationale for our model structure. Section 3 lays out the model and estimation procedure, and section 4 presents a numerical simulation to assess the accuracy of the model in recovering the true parameter values. An empirical application of the model is set forth in section 5 to study the interdependence in a binary choice decision: whether to buy a foreign brand or a domestic brand of mid-sized car. Section 6 offers concluding comments.

2. Review of Interdependent Preference Studies

People do not live in a world of isolation. They interact with each other when forming their opinions, beliefs and preferences. Interdependent preference (or preference interaction) has been defined as “occurring when an agent’s preference ordering over the alternatives in a choice set depends on the actions chosen by other agents” (Manski 2000). This effect has also been called "peer influences" (Duncan, Haller and Portes 1968), "neighborhood effects" (Case 1991), “bandwagon effect” (Leibenstein 1950) and "conformity" (Bernheim 1994) in studies of interdependence.
Interdependent preferences can arise in many ways. Reasons include social concerns, reductions in transaction costs (i.e., network externalities) and the signaling effect of another's brand ownership on inferred attribute-levels. In addition, interdependent preferences may appear to be present in economic demand models in which key explanatory variables are either omitted (e.g., household income) or unobserved (e.g., media exposure). Our review of the literature focuses on the former explanation, and the model specification is guided by the potential mechanisms by which interdependent preference arises. The review below provides a useful guide to researchers studying the origins of utility.

Pioneering work on interdependent preference was conducted by Duesenberry (1949) and Leibenstein (1950). Duesenberry described several examples of interdependence in consumer consumption behavior. Using data on consumer purchases made in 1935-36, he found that the percentage of income spent on consumption is highly correlated with the person’s rank order in the local income distribution. Leibenstein (1950) formally incorporated the phenomenon of "conspicuous consumption" into a theory of consumer demand. Through a conceptual experiment, he showed that the demand curve will be more elastic if there is a bandwagon effect than if the demand is based only on the functional attributes of the commodity. Since this work, a large body of literature has emerged that has examined the theory and empirical evidence of interdependent preference.

*Theoretical Research on Interdependent Preference*

In the theory domain, Hayakawa and Venieris (1977) derived several utility theories and axioms for preference interdependence. Their theories predict that the income effect associated with a price change will become dominant as the budget expenditure is relaxed. Moreover, their
research points to the need for a concept of psychological complementarity to capture the role of reference group in a consumer choice calculus. While Hayakawa and Venieris’s framework is mainly static, several other papers have tried to address the influence of preference interdependence adding a temporal dimension. For example, Cowan, Coman and Swann (1997) derived the steady state and dynamic properties of the distribution of consumption when different reference groups were used. Bernheim (1994) suggests a theory of how standards of behavior might evolve in response to changes in the distribution of intrinsic preferences.

While the focus of studying the interdependent effects in economics is mainly on its impact on demand theory and econometric implications, researchers in the marketing have paid more attention to explaining the mechanisms of the interdependent preferences among consumers in the context of reference group formation and influence. These two areas of research are distinguished. The first examines the reasons that individuals conform to the behavior of a reference group. Researchers have identified three sources of social influence on buyer behavior: internalization, identification and compliance (Kelman 1961; Burnkrant and Consineau 1975; Lessig and Park 1977). Internalization occurs when a person adopts other people’s influence because it is perceived to "inherently conductive to the maximization of his value." In other words, people are willing to learn from others in a sense that this could help them make a better decision that optimizes their own returns. Identification occurs when adopting from others because the "behavior is associated with a satisfying self-defining relationship" to the other. Compliance occurs when "the individual conforms to the expectations of another in order to receive a reward or avoid punishment mediated by that other."

The second area of research in marketing examines the relative influence of alternative mechanisms on individual consumer behavior. Bearden and Etzel (1982) studied how reference
groups influence an individual consumer’s purchase decisions at both product and brand level. Lessig and Park (1982) showed that the degree of reference group influence is dependent on the product related characteristics such as complexity, conspicuousness and brand distinction. Childers and Rao (1992) found that reference-group influence varies for products consumed in different occasions (public vs. private) and for different reference groups (familial vs. peer).

Interdependent preferences can therefore be associated with multiple covariates, and can lead to either conformity or individuality in preferences. In the next section we present a model capable of reflecting these effects, within the framework of an economic choice model.

*Measuring Interdependence*

In the empirical domain, a considerable amount of effort has been devoted to developing econometric models and estimation methods that incorporate interpersonal dependence (Pollak 1976; Kapteyn 1977; Stadt, Kapteyn and Geer 1985). Some empirical applications include studying interdependent preference in consumer expenditure allocations (Darough, Pollak and Wales 1983; Alessie and Kapteyn 1991; Kaptyen et. al. 1997), labor supply (Aronsson, Blomquist and Sacklen 1999), rice consumption (Case 1991), and elections (Smith and LeSage 2000). The focus of these studies is typically very aggregate, with the dependent variable reflecting average behavior (e.g. consumption) within a particular geographic region (e.g. zip code). Furthermore, they tend to use a single network to model the preference interdependence.

Methodological research in marketing on interdependent preferences has recently been spurred by developments in simulation-based estimation that facilitate flexible models of consumer heterogeneity, including models where dependence is spatially related. Arora and Allenby (1999) propose a conjoint model where the importance of product attributes in a group
decision making context can be different from the part-worths of each individual. However, their model assumes that the social group is readily identified, and does not account for the possibility of multiple networks. Hofstede, Wedel and Steenkamp (2002) examine the use of alternative spatial prior distributions to geographically smooth model parameters in a study of retail store attributes. In their analysis, response coefficients in a geographic area are assumed to be similar to neighboring areas. Bronnenberg and Mahajan (2001) combine an autoregressive spatial prior on market shares with a temporal autoregressive process to study variation in the effectiveness of promotional variables in geographically defined markets. Bronnenberg and Sismeiro (2002) use a spatial model to forecast brand sales in markets where only limited information exist. Neither of these later studies are developed to study preferences at the level of the individual consumer.

In this paper we develop an autoregressive mixture model and apply it to the latent utility in a discrete choice model. The autoregressive model relates an individual's latent utility to the utility of other individuals, reflecting the potential interdependence of preferences. Explanatory variables are incorporated into the autoregressive process through a weighting matrix that describes the network. The mixture aspect of the model allows the weighting matrix to be defined by multiple covariates, with covariate importance estimated from the data. Covariates are also related to the expected value of the latent utilities to capture exogenous, as opposed to endogenous effects.

3. A Hierarchical Bayes Autoregressive Mixture Model

In this section, we first introduce a binary choice model that captures the potential social dependency of preferences among individuals. Then we briefly describe the prior distribution
specification and estimation procedure using Markov chain Monte Carlo methods. A detailed description of the estimation algorithm is provided in the appendix.

3.1 An Autoregressive Discrete Choice Model

Suppose we observe choice information for a set of individuals \( i = 1, \ldots, m \) who are not associated with an interdependent network and whose preferences are exogenously determined. Assume that the individual is observed to make a selection between two choice alternatives \( y_i = 1 \) or \( 0 \) that is driven by the difference in latent utilities, \( U_{ik} \), for the two alternatives \( k = 1, 2 \). The probability of selecting the second alternative over the first is:

\[
\Pr(y_i = 1) = \Pr(U_{i2} > U_{i1}) = \Pr(z_i > 0)
\]

(1)

\[
z_i = x_i' \beta + \varepsilon_i
\]

(2)

\[
\varepsilon_i \sim \text{Normal}(0,1)
\]

(3)

where \( z_i \) is the latent preference for the second alternative over the first alternative, \( x_i \) is a vector of covariates that captures the differences of the characteristics between the two choice alternatives and characteristics of the individual, \( \beta \) is the vector of coefficients associated with \( x_i \), and \( \varepsilon_i \) reflects unobservable factors modeled as error. Preference for a durable offering, for example, may be dependent on the existence of local retailers who can provide repair service when needed. This type of exogenous preference dependence is well represented by equation (2). The error term is assumed to be independently distributed across individuals, reflecting the absence of interdependent effects. The scale of the error term is equal to one to statistically identify the
model coefficients, \( \beta \). Stacking the latent preferences, \( z_i \), into a vector results in a multivariate specification:

\[
z \sim \text{Normal}(X\beta, I) . \tag{4}
\]

The presence of interdependent networks creates preferences that are endogenous and mutually dependent, resulting in an error covariance matrix (\( \Sigma \)) with non-zero off diagonal elements. The presence of off-diagonal elements in the covariance matrix leads to conditional and unconditional expectations of preferences that differ. The expectation of latent preference, \( z \), in equation (4) is equal to \( X\beta \) regardless of whether preferences of other individuals are known. However, if the off-diagonal elements of the covariance matrix are non-zero, then the conditional expectation of the latent preference for one individual is correlated with the revealed preference of another individual:

\[
E[z_2|z_1] = X_2'\beta + \Sigma_{21}\Sigma_{11}^{-1}(z_1-X_1\beta) , \tag{5}
\]

where the subscripts apply to appropriate elements of the parameters. Positive covariance leads to a greater expectation of preference \( z_2 \) if it is known that \( z_1 \) is greater than its mean, \( X_1\beta \). In the probit model, choice \( (y_i) \) is revealed, corresponding to a range of latent preferences (i.e., \( z_i > 0 \)). The computation of conditional expectation is therefore associated with an integration over a range of possible conditioning arguments.

An approach to inducing covariation among the error terms is to augment the error term in equation (2) with a second error term from an autoregressive process (LeSage 2000):
\[ z_i = x_i' \beta + \varepsilon + \theta_i \]  \hspace{1cm} (6)

\[ \theta = \rho W \theta + u \]  \hspace{1cm} (7)

\[ \varepsilon \sim N(0, I) \]  \hspace{1cm} (8)

\[ u \sim N(0, \sigma^2 I) \]  \hspace{1cm} (9)

where \( \varepsilon \) and \( u \) are iid error terms, \( \theta \) is a vector of autoregressive parameters where the matrix \( \rho W \) reflects the interdependence of preferences across individuals. The specification in equation (7) is similar to that encountered in time series analysis, except that co-dependence can exist between two elements, whereas in time series analysis the dependence is directional (e.g., an observation at time \( t-k \) can affect an observation at time \( t \), but not vice versa). Co-dependence is captured by non-zero entries appearing in both the upper and lower triangular sub-matrices of \( W \). It is assumed that the diagonal elements \( w_{ii} \) are equal to zero and each row sums to one. The coefficient \( \rho \) measures the degree of overall association among the units of analysis beyond that captured by the covariates, \( X \). Positive (negative) value of \( \rho \) indicates positive (negative) correlation among people.

It is worth noting that in our model of interdependent preferences, there is a network propagation effect captured in equation (7), where in the exogenous model the effect associated with a covariate does not propagate among consumers. Our model presents a simple test on the existence of propagation effect. If \( \rho \) is significantly different from zero, then we conclude that there could be some interdependent preference beyond what is captured in the \( x' \beta \) term in equation (6).

The augmented-error model results in latent preferences with non-zero covariance:
This specification\(^1\) is different from that encountered in standard spatial data models (see Cressie 1991, p.441) where the error term \(\varepsilon\) is not present and the covariance term is equal to \(\sigma^2(I-\rho W)^{-1}(I-\rho W')^{-1}\). The advantage of specifying the error in two parts is that it leads itself to estimation and analysis using the method of data augmentation (Tanner and Wong, 1987). The autoregressive parameter \(\theta\) is responsible for the nonzero covariances in the latent preferences, \(z\), but is not present in the likelihood specification (equation (10)). By augmenting the parameters space with \(\theta\), we isolate the effects of the non-zero covariances and simplify the evaluation of the likelihood function (see below).

The elements of the autoregressive matrix, \(W = \{w_{ij}\}\), reflect the potential dependence between units of analysis. A critical part of the autoregressive specification concerns the construction of \(W\). Spatial models, for example, have employed a coding scheme where the un-normalized elements of the autoregressive matrix equal one if \(i\) and \(j\) are neighbors and zero otherwise (see Bronnenberg and Mahajan 2001). An alternative specification for a spatial model could involve other metrics, such as Euclidean and Manhattan distances. However, as noted above, interdependent preferences can be determined by multiple networks. It is therefore important to allow for a specification of the autoregressive matrix, \(W\), with multiple covariates.

\(^1\) With this model specification, we assume that \(\theta\) has smaller variance for people in larger networks. We believe this property is justified. As the size of the network increases, interdependence implies there exists more information about a specific consumer's preference, and hence smaller variance.
We specify the autoregressive matrix $W$ as a finite mixture of coefficient matrices, each related to a specific covariate:

$$W = \sum_{k=1}^{K} \phi_k W_k,$$  \hspace{1cm} (11)

$$\sum_{k=1}^{K} \phi_k = 1. \hspace{1cm} (12)$$

where $k$ indexes the covariates, $k = 1, \ldots, K$. The weights, $\phi_k$, reflect the relative importance of the component matrices, $W_k$, with each associated with a different explanatory variable. $W_1$, for example, may be related to the physical proximity of the individual residences, $W_2$ to their age, $W_3$ to their income, $W_4$ to their ethnicity, and so on. Within each matrix, $W_k$, the diagonal elements are assumed equal to zero, the off diagonal elements reflect the distance between individuals in terms of the $k$th covariate, and each row sums to one. The weighted sum of them component matrices, $W$, also has these properties since the weights, $\phi_k$, sum to one. In our specification we re-parameterize $\phi_k$ with a logit specification:

$$\phi_j = \frac{\exp(\alpha_j)}{\sum_{k=1}^{K} \exp(\alpha_k)} \hspace{1cm} (13)$$

and estimate $\alpha_j$ unrestricted with $\alpha_K = 0$.

The model described by equations (1), (6), (7), (8) and (9) is statistically identified. This can be seen by considering the choice probability $\Pr(y_i = 1) = \Pr(z_i > 0) = \Pr(x_i'\beta + \varepsilon_i + \theta_i > 0)$.
The right side of the later expression is zero, and the variance of $\varepsilon_i$ is one. These specifications identify the probit model in terms of location and scale because an arbitrary constant cannot be added to the right side of the expression, and multiplying by a scalar quantity would alter the variance of $\varepsilon_i$. Moreover, the finite mixture specification described by equation (10), (11) and (12) is identified because the rows of $W_k$ and the mixture probabilities are normalized to add to one. However, we note that care must be exercised in comparing estimates of $\beta$ from the independent probit model (equation (4)) with that from the autoregressive model (equation (10)) because of the differences in the magnitude of the covariance matrix. We return to the issue of statistical identification below when demonstrating properties of the model in a simulation study.

Our model is based on the framework developed by Smith and LeSage (2001), but extends theirs in the following ways. First, the spatial matrix $W$ in their model is constrained to be only geographic specific. We introduce a more flexible structure to capture different sources of interdependence across multiple networks. Second, most applications in this area of research, including Smith and LeSage (2001), focus on zip code level analysis. Our analysis investigates interdependence at the individual level which is of interest to marketing.

3.2 Prior Specification and Markov chain Estimation

We estimate the autoregressive discrete choice model using Markov chain Monte Carlo methods. This method of estimation requires specification of prior distributions for the model parameters in equation (10), and derivation of the full conditional distribution of model parameters. The prior distributions are set to be diffuse and conjugate when possible. We use some standard prior distribution specification as follows:
\[ \beta \sim N(\beta_0, V_\beta) \]  
\[ \rho \sim U\left[\frac{1}{\lambda_{\min}}, \frac{1}{\lambda_{\max}}\right] \]  
\[ \alpha \sim N(\alpha_0, V_\alpha) \]  
\[ \sigma^2 \sim IG(s_0, q_0) \]

Here, \( \alpha \) and \( \beta \) have normal conjugate prior distributions with means set to zero and covariance matrices set to 100I where I is the identity matrix, and \( \sigma^2 \) is assigned a conjugate inverted gamma prior with \( s_0 = 5 \) and \( q_0 = 0.1 \). We employ a uniform prior distribution on \( \rho \) over a specified range. The parameter \( \rho \) must lie in this interval \( \left[\frac{1}{\lambda_{\min}}, \frac{1}{\lambda_{\max}}\right] \), where \( \lambda_{\min} \) and \( \lambda_{\max} \) denote the minimum and maximum eigenvalues of \( W \), for the matrix \( (I - \rho W) \) to be invertable (Sun, Tsutakawa and Speckman 1999).

The Markov chain proceeds by generating draws from the set of conditional posterior distributions of the parameters. As mentioned above, we augment the model parameters \( \theta \) in equation (7) that capture the dependent error structure through an autoregressive process. By conditioning on \( \theta_i \), the latent preference, \( z_i \), is seen to arise from a standard binomial probit model with mean \( x_i'\beta + \theta_i \) and independent errors. Furthermore, the conditional distributions of the model parameters, given \( \theta \), are of standard form. A detailed description of the full conditional distributions is provided in Appendix A. We note that generating draws from the full conditional distribution of \( \theta \) is computationally demanding, and we adopted the method proposed
by Smith and LeSage (2000) to iteratively generate draws for the elements of $\theta$. Appendix B briefly outlines this method.

4. A Simulation Study

In this section, we demonstrate properties of the autoregressive choice model and investigate the relationship between sample size and accuracy of the MCMC estimator. We focus our analysis on the model without the latent mixing distribution described in equation (11) through equation (13).

4.1 Data Simulation

We simulate three datasets. The first dataset is comprised of 50 individuals and the second dataset contains 500 individuals. We assume the individuals are connected circularly, with each individual affected by his/her two closest neighbors. To illustrate, the autoregressive matrix, $W$, for five individuals is as follows:

$$
W = \begin{bmatrix}
0 & .5 & 0 & 0 & .5 \\
.5 & 0 & .5 & 0 & 0 \\
0 & .5 & 0 & .5 & 0 \\
0 & 0 & .5 & 0 & .5 \\
.5 & 0 & 0 & .5 & 0
\end{bmatrix}
$$

We assume that $X$ is comprised of two covariates generated from a standard normal distribution, $\beta' = (1,1)'$, $\sigma^2 = 4$ and $\rho = 0.5$. Binary choices ($y_i$) are simulated by generating draws from a multivariate normal distribution specified in equation (10) and applying the censoring described
in equation (1). The covariance matrix of the latent preference distribution for the five individuals described by the autoregressive matrix in equation (18) is:

\[
\Sigma = I + \sigma^2 (I - \rho W)^{-1} (I - \rho W')^{-1} = \begin{bmatrix}
7.3 & 3.2 & 1.6 & 1.6 & 3.2 \\
3.2 & 7.3 & 3.2 & 1.6 & 1.6 \\
1.6 & 3.2 & 7.3 & 3.2 & 1.6 \\
1.6 & 1.6 & 3.2 & 7.3 & 3.2 \\
3.2 & 1.6 & 1.6 & 3.2 & 7.3
\end{bmatrix}
\] (19)

The covariance between the first and third individual is nonzero despite the fact that these two individuals are not neighbors. The first individual is connected to the second individual and the second to the third in the autoregressive matrix (equation (18)). The connections induce nonzero covariance between the first and third individuals, reflecting the correlation of the circular network. The variances along the diagonal of \(\Sigma\) are equal, reflecting the fact that each individual has exactly two neighbors. To examine the performance of the estimator for a heteroskedastic error matrix, we include a third case in which the matrix \(W\) is formulated using data from our empirical study (reported below). This third simulation study contains 666 observations.

4.2 Accuracy of Parameter Estimates

Estimation results are presented in Table 1. We ran the Markov chain for 5000 iterations and deleted the first 1000 draws for 'burn-in' of the chain. The last 4000 draws were used to calculate the posterior mean and standard deviation of the parameters. From the table, we see that the coefficient estimates are close to their true values, even in small samples. As expected, the true parameters lie inside the 95% highest posterior density intervals of the posterior
distributions, and the accuracy of the estimator improves as the sample size increases. We also ran the Markov chain for 10,000 iterations and found no appreciable difference in the estimates reported in table 1. These results support our conclusion that the autoregressive choice model is statistically identified and our estimation using data augmentation is valid.

\[ \text{Table 1} \]

The purpose of the autoregressive specification is to understand dependence among individuals as reflected in the covariance matrix. It is therefore important to investigate the model's ability to recover the realizations of the choice model error responsible for inducing the covariances. Figure 1 provides a comparison between the estimated and actual autoregressive effects \( \theta_i \). The top panel of figure 1 displays a plot of the estimated and actual autoregressive effects for the dataset with 50 observations (individuals), the middle panel of figure 1 displays the plot for the dataset with 500 observations, and the bottom panel of figure 1 displays the plot for the dataset with 666 observations with spatial interaction generated to mimic our empirical analysis, described below. Also included in the plots is a 45 degree line. The points in the graphs tend to fall evenly about the line, indicating that the estimated autoregressive effects are not biased. Moreover, the variability of the points about the line does not depend appreciably on the sample size. An increase in the sample size results in improved estimates of the model parameters in table 2, but not in terms of the augmented parameter \( \theta_i \), because the dimension of \( \theta \) is the same as the number of observations in the analysis. Our analysis indicates that, despite the sparseness present in a circularly connected population of 500 people, or present in our data reported below, accurate estimates of model parameters, including the augmented parameters \( \theta \), are possible.
The autoregressive effects are used to assess the degree of dependence in the probit error structure. If there is little interdependence in consumer choice, or if the data is not sufficiently informative about the presence of interdependencies, then the realizations of \( \theta \) will be near zero and the predicted choice probabilities, conditional on \( \theta \), will be similar to that obtained from an independent probit specification \((\Sigma = I)\). If choice and preferences are interdependent, then the realizations of \( \theta \) will be large and there will exist improvements in model fit by conditioning on the dependent information contained in \( \theta \).

4.3 Assessing Differences in Regression Coefficients

The presence of an autoregressive component in the error term changes the variance of the probit model from the identity matrix to \( \Sigma = I + \sigma^2(I-\rho W)^{-1}(I-\rho W')^{-1} \). As illustrated in equation (19), the autoregressive specification can lead to large changes in the diagonal elements of the covariance matrix. Care must therefore be exercised when interpreting the regression coefficients associated with the mean of the multivariate normal distribution. Although we have demonstrated that the autoregressive specification leads to an identified model, these coefficients must be interpreted relative to the scale of the error term. This is most easily accomplished by computing the expected change in the choice probability for a change in the independent variable.

Table 2 contains the regression coefficient estimates for the dataset with 500 observations discussed above, along with regression estimates from a traditional binary probit model with \( \Sigma = I \) (that is, setting all \( \theta \) to 0). The regression estimates for the autoregressive model are much larger than those obtained from the independent probit model. However, when these coefficients
are converted to expected derivatives of the choice probability at $x_1 = 0$ and $x_2 = 0$, the estimates for the two models are seen to closely agree. The change in probability is 0.245 for the independent probit model (equation (4)), and 0.270 for the dependent probit model (equation (10)) when both $x_1$ and $x_2$ increase from 0.0 to 1.0. Either model is capable of capturing the average association between the covariates and choice. The autoregressive model, however, is needed to understand the extent and nature of preference inter-dependencies given the average association.

\[= = \text{Table 2} = =\]

5. Empirical Application

Data were collected by a marketing research company on purchases of mid-sized cars in the United States. We obscure the identity of the cars for the purpose of confidentiality. The cars are functionally substitutable, priced in a similar range, and are distinguished primarily by their national origin: Japanese and non-Japanese. Japanese cars have a reputation for reliability and quality in the last twenty years, and we seek to understand the extent to which preferences are inter-dependent among consumers.

We investigate two sources of dependence – geographic and demographic neighbors. Geographic neighbors are created by physical proximity and measured in terms of geographic distance among individuals' places of residence. Demographic neighbors are defined in terms of similar demographic variables. Young people, for example, are more likely to associate with other young people, obtain information from them, and may want to conform to the beliefs to their reference group to gain group acceptance and social identity. We empirically test these referencing structures and analyze their importance in driving the preferences.
We operationalize the different referencing schemes as follows. The data include information on the longitude and latitude information of each person’s residence, and we can calculate the geographic distance between person i and j as follows:

\[ d(i, j) = \sqrt{(d^1_i - d^1_j)^2 + (d^2_i - d^2_j)^2} \]  

(20)

where \( d^1_j \) denotes the longitude and \( d^2_j \) denotes the latitude of individual j’s home. We further assume that geographic influence is an inverse function of the geographic distance:

\[ w_{i,j}^{\text{geo}} = \frac{1}{\exp(d(i, j))} \]  

(21)

An alternative geographic specification of \( W \) that leads to a symmetric matrix is to identify neighbors by the zip code of their home mailing address:

\[ w_{i,j}^{\text{zip}} = \begin{cases} 1 & \text{if person i and person j have the same zip code;} \\ 0 & \text{otherwise.} \end{cases} \]  

(22)

We operationalize demographic neighbors in terms of people who share characteristics such as education, age, income, etc. Individuals in the dataset are divided into groups defined by age of the head of the household (3 categories), annual household income (3 categories), ethnic affiliation (2 categories) and education (2 categories). This leads us to a maximum of 36 groups, 31 of which are presented in our sample. The demographic specification of \( W \) becomes:
\( w_{i,j}^d = 1 \) if person \( i \) and person \( j \) belong to the same demographic group; 0 otherwise.

(23)

The data consist of 857 consumers who live in 122 different zip codes. Table 3 provides sample statistics of the data. Approximately 85% of people purchased a Japanese car. On average, the price a Japanese car is $2400 cheaper than a non-Japanese car and there is little difference between the optional accessories purchased with the cars. We note, however, that the sample standard deviations are nonzero, indicating intra-group variation. The average age of the consumer is 49 years, and average annual household income is approximately $67,000. Approximately 12% of the consumers are of Asian origin, and 35% have earned a college degree. The choices of 666 individuals from 100 zip codes are used to calibrate the model, and 191 observations form a holdout sample. Figure 2 is a histogram of the number of zip codes containing at least two consumers, indicating that the sample of respondents is geographically dispersed.

Table 3

Figure 2

In sample and out of sample fit are assessed in the following way. When possible, we report fit statistics conditional on the augmented parameter \( \theta_i \). For the independent probit model (equations (1) – (3)) in which \( \theta_i = 0 \), the latent preferences are independent after accounting for the influence of the covariates in mean of the latent distribution (equation (4)). Knowledge that an individual actually purchased a Japanese car provides no help in predicting the preferences and choices of other individuals. When preferences are interdependent, information about others'
choices is useful in predicting choices, and this information is provided through $\theta_i$ as described in equation (7).

In-sample fit is assessed using the importance sampling method of Newton and Raftery (1994, p.21) that re-weights the conditional likelihood of the data. Conditional on $\theta$, this evaluation involves the product of independent probit probabilities and is easy to compute. Computing the out of sample fit is more complicated. Our analysis proceeds as it would in developing a customer scoring model, by constructing the autocorrelation matrices $W$ for the entire dataset (857 observations), but estimating the model parameters using the first 666 observations. We obtain the augmented parameters for the holdout sample, $\theta^p$, by noting that

$$
\begin{pmatrix}
\theta \\
\theta^p
\end{pmatrix} 
\approx N \left( 0, \sigma^2 (I_{857} - \hat{\rho} \hat{W})^{-1} (I_{857} - \hat{\rho} \hat{W})^{-1} \right) = N \left( 0, \begin{pmatrix} \Sigma_{11} & \Sigma_{12} \\ \Sigma_{21} & \Sigma_{22} \end{pmatrix} \right) \tag{24}
$$

where $\Sigma_{12}$ is the covariance matrix between $\theta$ and $\theta^p$, and we can obtain the conditional distribution of $\theta^p$ given $\theta$ using properties of the multivariate normal distribution:

$$
\left( \theta^p \mid \theta \right) \sim MN(\mu, \Omega) \tag{25}
$$

where

$$
\mu = \Sigma_{21} \Sigma_{11}^{-1} \theta \tag{26}
$$

$$
\Omega = \Sigma_{22} - \Sigma_{21} \Sigma_{11}^{-1} \Sigma_{12} \tag{27}
$$
Table 4 reports the in sample and out sample fit statistics for six different models. The first model is an independent binary probit model (equations 1-3) where the probability of purchasing a Japanese car is associated with the feature differences between cars, demographics information for the individual, geographic information (longitude, latitude), and dummy variables for the demographic groups. The dummy variables can be viewed as an attempt to capture high-order interactions of the covariates. Dummy variables for only 19 of the 31 groups are included because the proportion of buyers of Japanese cars in the remaining groups is at or near 100%. Thus, the first model attempts to approximate the structure of heterogeneity using a flexible exogenous specification.

The second model is a random-effects model that assumes people living in the same zip code have identical price and option coefficients. Geographic and demographic variables are incorporated into the model specification to adjust the model intercept and the mean of the random-effects distribution. The second model represents a standard approach to modeling preferences, incorporating observed and unobserved heterogeneity.

Models 3-6 specify four variations of interdependent models in which consumer preferences are endogenous, or interdependent. Models 3 and 4 are alternative specifications for geographic neighbors (equations (21) and (22)), whereas Model 5 specifies the autoregressive matrix in terms of the 31 demographic groups (equation 23). Model 6 incorporates both geographic and demographic structures using the finite mixture model in equations (11) – (13).

The model fit statistics (both in sample and out sample) indicate the following. First, car choices are inter-dependent. Model 1 is the worst fitting model and all attempts into incorporate geographic and/or demographic interdependence in the model leads to improved in-sample and
out-of-sample fit. A comparison of the fit statistics between Model 2 and Models 3-6 indicates that there is stronger evidence of interdependence in the autoregressive models than in a random-effects model based on zip codes. Moreover, adding quadratic and cubic terms for longitude and latitude results in a slight improvement in-sample fit (i.e., from –203.809 to –193.692) and an out-of-sample fit (i.e., from 0.158 to 0.154). This supports the view that people have similar preferences not only because they share similar demographic characteristics that may point to similar patterns of resource allocation (an exogenous explanation), but also because there is an endogenous interdependence among people. Furthermore, Model 3 and 4 produce very similar fit statistics, which indicates that geographic neighbor based weighting matrix performs a good approximation to geographic distance based weighting matrix. The introduction of both of geographic and demographic referencing schemes in Model 6 leads to an improvement in the fit, showing that both reference groups are important in influencing the individual’s preference.

Parameter estimates for the six models are reported in table 5. We note that, in general, the coefficient estimates are largely consistent across the six models, indicating that all of the models are somewhat successful in reflecting the data structure. Since Model 6 yields the best in-sample and out-sample fit, we focus our discussion on its parameter estimates. The estimates indicate that price, age, income, ethnic origin, longitude and latitude are significantly associated with car purchases, with Asians, younger people, people with high incomes, and living more southern and more western preferring Japanese makes of car. The ethnic variable turns out to have a very large coefficient indicating that Asian people will have a significantly higher likelihood of choosing a Japanese brand car. Furthermore, $\rho$ is significantly positive indicating a positive correlation among consumer preferences. $\alpha$ is significantly positive ($\phi$ greater than
indicating that geographic reference groups are more important in determining the individual’s preference compared with demographic reference groups.

Figure 3 displays estimates of the elements of the augmented parameter $\theta$ against the longitude and latitude of each observation in sample. Most of the estimates have posterior distributions away from zero, providing evidence of interdependent choices. An analysis of the difference in covariates for the two groups ($\theta > 0$ and $\theta \leq 0$) does not reveal any statistically significant differences except for the longitude variable. That is, the augmented parameters that capture the endogenous nature of preferences are not simply associated with all the covariates in the analysis.

6. Summary and Concluding Remarks

In this paper we introduce an autoregressive multivariate binomial probit model to study interdependent choices among individuals. The model is specified in a hierarchical Bayes framework and estimation algorithms are derived using data augmentation to simplify the computations. We investigate the influence of two possible sources of interdependent influence: geographic neighbors and demographic neighbors. Geographic neighbors are individuals who reside in close proximity to each other, while demographic neighbors are identified by demographic variables that point to social networks.

Alternative model specifications are used to investigate variation in preferences. We find that a standard random-effect specification is inferior to an autoregressive specification. In the random-effect specification, variation in preferences across individuals is modeled as
independent draws from a mixing distribution, whereas in an autoregressive specification the variation in preferences propagates through the networks. If person i is a neighbor of person j, and j is a neighbor of k, then person i can influence person k through person j. Such dependencies are not well reflected in iid draws from a mixing distribution.

We apply the autoregressive model to a dataset where the dependent variable is whether an individual purchases a Japanese make of car. Our empirical application demonstrates that: (i) there is a preference interdependence among individual consumers that reflects conformity ($\rho > 0$); (ii) the preference interdependence is more likely to take an endogenous influence structure than a simple exogenous structure; (iii) the geographically defined network is more important in explaining individual consumer behavior than demographic network. However, since our data is cross-sectional, we are unable to identify the true cause of the interdependence.

Choice models have been used extensively in the analysis of marketing data. In these applications, most of the analysis depends on the assumption that individual forms his or her own preferences and makes a choice decision irrespective of other people’s preferences. However, people live in a world where they are inter-connected, information is shared, recommendations made, and social acceptance is important. Interdependence is therefore a more realistic assumption in models of preference heterogeneity.

Our model can be applied and extended in many ways. Opinion leaders, for example, are individuals that exert a high degree of influence on others, and could be identified with extreme realizations of the augmented parameter, $\theta_i$. Aspiration groups that affect others, but are not themselves affected, could be modeled with an autoregressive matrix $W$ that is asymmetric. Temporal aspects of influence, including the word of mouth and 'buzz' (Rosen 2000) could be investigated with longitudinal and cross-sectional data with the autoregressive matrix $W$ defined
on both dimensions. Such time series data would help us to identify the source and nature of
interdependence. Finally, the model can be extended to apply to multinomial response data to
investigate the extent of interdependent preference in brand purchase behavior, or extended to
study the interdependence in $\beta$ coefficients across people. These applications and extensions
will contribute to our understanding of extended product offerings and the appropriateness of
"iid" heterogeneity assumptions commonly made in models of consumer behavior.
References


Appendix A: Markov chain Monte Carlo Estimation

Estimation is carried out by sequentially generating draws from the following distributions:

1. Given the choice, a latent continuous variable $z$ can be generated for the probit model.

   Generate $\{z_i, i=1,\ldots,m\}$

   $$f(z_i \mid *) = \text{Truncated Normal} \left( x_i \beta + \theta, 1 \right)$$

   if $y_i = 1$ then $z_i \geq 0$

   if $y_i = 0$ then $z_i < 0$

2. Generate $\beta$

   $$f(\beta \mid *) = MN(v, \Omega)$$

   $$v = \Omega(X'(z - \theta) + D^{-1} \beta_0)$$

   $$\Omega = (D^{-1} + X'X)^{-1}$$

   $$\beta_0 = (0,0,\ldots,0)'$$

   $$D = 400I_m$$

3. Generate $\theta$

   $$f(\theta \mid *) = MN(v, \Omega)$$

   $$v = \Omega(z - X\beta)$$

   $$\Omega = (D^{-1} + \sigma^{-2} B'B)^{-1}$$

   $$B = I - \rho W$$
4. Generate $\sigma^2_y$

$$f(\sigma^2_y | \gamma^*) \propto \text{Inverted Gamma}(a,b)$$

$$a = s_o + m/2 \quad (s_o = 5)$$

$$b = \frac{2}{\theta'B'\theta + 2/q_o} \quad (q_o = 0.1)$$

5. Generate $\rho$

We use Metropolis-Hastings algorithm with a random walk chain to generate draws (see Chib and Greenberg 1995). Let $\rho^{(p)}$ denote the previous draw, and then the next draw $\rho^{(n)}$ is given by:

$$\rho^{(n)} = \rho^{(p)} + \Delta$$

with the accepting probability $\alpha$ given by:

$$\min \left[ \frac{\left| B(\rho^{(n)}) \right| \exp \left\{ -0.5(1/\sigma^2)\theta'B(\rho^{(n)})'B(\rho^{(n)})\theta \right\} }{\left| B(\rho^{(p)}) \right| \exp \left\{ -0.5(1/\sigma^2)\theta'B(\rho^{(p)})'B(\rho^{(p)})\theta \right\} } \right]$$

$\Delta$ is a draw from the density Normal(0, 0.005). The choice for parameters of this density ensures an acceptance rate of over 50%. If $\rho$ lies outside the range of $\left[ \frac{1}{\lambda_{\min}}, \frac{1}{\lambda_{\max}} \right]$, the likelihood is assumed zero and we reject the candidate $\rho^{(p)}$.

6. Generate $\alpha$

$$W = \sum_{k=1}^{K} \phi_k W_k$$

and

$$\phi_j = \frac{\exp(\alpha_j)}{\sum_{k=1}^{K} \exp(\alpha_k)}$$
We elect to estimate \( \alpha \) rather than \( \phi \) directly. We use Metropolis-Hastings algorithm with a random walk chain to generate draws (similar to generating \( \rho \)). Let \( \alpha^{(p)} \) denote the previous draw, and then the next draw \( \alpha^{(n)} \) is given by:

\[
\alpha^{(n)} = \alpha^{(p)} + \Delta
\]

with the accepting probability \( \alpha \) given by:

\[
\min \left[ \frac{|B(\alpha^{(n)})| \exp \left\{ -0.5(1/\sigma^2) \theta' B(\alpha^{(n)})' B(\alpha^{(n)}) \theta \right\} \exp \left\{ -0.5(\alpha^{(n)} - \alpha_0)' T_0^{-1} (\alpha^{(n)} - \alpha_0) \right\}}{|B(\alpha^{(p)})| \exp \left\{ -0.5(1/\sigma^2) \theta' B(\alpha^{(p)})' B(\alpha^{(p)}) \theta \right\} \exp \left\{ -0.5(\alpha^{(p)} - \alpha_0)' T_0^{-1} (\alpha^{(p)} - \alpha_0) \right\}} \right]^{1/2}
\]

\( \Delta \) is a draw from the density \( \text{Normal}(0, 0.005I) \). \( \alpha_0 \) is a vector of 0, and \( T_0 \) is a prior covariance matrix with diagonal elements being equal to 100 and 0 for all off-diagonal elements. The choice for parameters of this density ensures an acceptance rate of over 50%.
Appendix B: An Approach to Efficiently Generate $\theta$

In order to generate $\theta$, we need to invert a $m \times m$ matrix $(D^{-1} + \sigma^{-2}B'B)$. When $m$ is large, it is computational burdensome to make the inversion. We adopt an efficient method of generating $\theta$ from its posterior distribution while avoiding inverting a high dimensional matrix introduced by Smith and LeSage (2000). The essence is to generate univariate conditional posteriors of each component $\theta_i$ rather than a joint posterior distribution of the vector of $\theta$.

Next, we briefly describe the procedure.

$$\theta \propto \exp\{A\}$$

where

$$A = \exp\left\{ -0.5[\theta'(\sigma^{-2}B'B + I^{-1})\theta - 2(z - X\beta)'\theta] \right\}$$

$$= \sigma^{-2} \{\theta'\theta - 2\rho\theta'W\theta + \rho^2\theta'W'W\theta\} + \theta'\theta - 2\phi'\theta$$

where

$$\phi = (\phi_i = z_i - x_i'\beta : i = 1,...,m)'$$

Further decompose the vector of $\theta$ as $(\theta_i, \theta_{-i})$ and define $w_j$ to be the i’th column of $W$ and $W_{-i}$ to be a $m \times (m-1)$ matrix of all other columns of $W$. Then we obtain the following expressions:

$$\theta'W\theta = \theta_i'w_i + \sum_{j \neq i} \theta_j (w_j + w_{-i}) + C$$

$$\theta'W'W\theta = \theta_i^2w_i'w_j + 2\theta_i (w_j'W_{-i}\theta_{-i}) + C$$

$$\theta'\theta = \theta_i^2 + C$$

$$-2\phi'\theta = -2\phi_i\theta_i + C$$

where $C$ denotes a constant that does not involve parameters of interest. Substituting the above expressions back and we can write the conditional posterior distribution of $\theta_i$ as follows:
\[ f(\theta_i | *) \propto \exp\{-0.5(a_i \sigma_i^2 - 2b_i \theta_i)\} = N(b_i / a_i, 1 / a_i) \]

where

\[ a_i = 1/\sigma^2 + \rho^2 / \sigma^2 w_j'w_i + 1 \]

\[ b_i = \phi_i + \rho / \sigma^2 \sum_{j \neq i} \theta_j (w_{ji} + w_{ji}) - \rho^2 / \sigma^2 w_j'W_{-i} \theta_{-i} \]

We compared speed of this algorithm relative to generating \( \theta \) by directly inverting the matrix \((D^{-1} + \sigma^{-2}B' B)\), and sampling from a multivariate distribution. The algorithm results in a 5-fold decrease in the total time required for one iteration of the 6 step algorithm for the second simulation study involving 500 observations.
Table 1. Estimates from the Numerical Simulation

<table>
<thead>
<tr>
<th>N</th>
<th>W</th>
<th>True Values</th>
<th>$\beta_1$</th>
<th>$\beta_2$</th>
<th>$\rho$</th>
<th>$\sigma^2$</th>
</tr>
</thead>
<tbody>
<tr>
<td>50</td>
<td>from simulation</td>
<td>Posterior Mean</td>
<td>1.286</td>
<td>0.884</td>
<td>0.608</td>
<td>3.372</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Posterior Stdev.</td>
<td>(0.622)</td>
<td>(0.501)</td>
<td>(0.089)</td>
<td>(1.859)</td>
</tr>
<tr>
<td>500</td>
<td>from simulation</td>
<td>Posterior Mean</td>
<td>0.951</td>
<td>0.891</td>
<td>0.510</td>
<td>4.010</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Posterior Stdev.</td>
<td>(0.281)</td>
<td>(0.290)</td>
<td>(0.061)</td>
<td>(1.944)</td>
</tr>
<tr>
<td>666</td>
<td>from real data</td>
<td>Posterior Mean</td>
<td>0.921</td>
<td>0.946</td>
<td>0.474</td>
<td>3.250</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Posterior Stdev.</td>
<td>(0.153)</td>
<td>(0.163)</td>
<td>(0.058)</td>
<td>(1.215)</td>
</tr>
</tbody>
</table>
Table 2
Differences in Regression Coefficients

<table>
<thead>
<tr>
<th>Model</th>
<th>True Values</th>
<th>$\beta_1$</th>
<th>$\beta_2$</th>
<th>$\rho$</th>
<th>$\sigma^2$</th>
</tr>
</thead>
<tbody>
<tr>
<td>Independent (Equation 4)</td>
<td></td>
<td>1</td>
<td>1</td>
<td>0.5</td>
<td>4</td>
</tr>
<tr>
<td>Posterior Mean</td>
<td>0.350</td>
<td>0.317</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>Posterior Stdev.</td>
<td>(0.058)</td>
<td>(0.064)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Dependent (Equation 10)</td>
<td></td>
<td>0.951</td>
<td>0.891</td>
<td>0.510</td>
<td>4.01</td>
</tr>
<tr>
<td>Posterior Mean</td>
<td>0.281</td>
<td>0.290</td>
<td></td>
<td></td>
<td>(1.644)</td>
</tr>
<tr>
<td>Posterior Stdev.</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
Table 3.
Sample Statistics

<table>
<thead>
<tr>
<th>Variable</th>
<th>Mean</th>
<th>Standard Deviation</th>
</tr>
</thead>
<tbody>
<tr>
<td>Car Choice (1=Japanese, 0=Non-Japanese)</td>
<td>0.856</td>
<td>0.351</td>
</tr>
<tr>
<td>Difference Price (in 100 $s)</td>
<td>-2.422</td>
<td>2.998</td>
</tr>
<tr>
<td>Difference in Options (in 100 $s)</td>
<td>0.038</td>
<td>0.342</td>
</tr>
<tr>
<td>Age of Buyer (in number of years)</td>
<td>48.762</td>
<td>13.856</td>
</tr>
<tr>
<td>Annual Income of Buyer (in 1000 $s)</td>
<td>66.906</td>
<td>25.928</td>
</tr>
<tr>
<td>Ethnic Origin (1=Asian, 0=Non-Asian)</td>
<td>0.117</td>
<td>0.321</td>
</tr>
<tr>
<td>Education (1=College, 0=Below College)</td>
<td>0.349</td>
<td>0.477</td>
</tr>
<tr>
<td>Latitude (relative to 30 = original -30)</td>
<td>3.968</td>
<td>0.484</td>
</tr>
<tr>
<td>Longitude (relative to –110 = original +110)</td>
<td>-8.071</td>
<td>0.503</td>
</tr>
</tbody>
</table>
Table 4.
Model Fit Comparison

<table>
<thead>
<tr>
<th>Model</th>
<th>Specification (a)</th>
<th>In Sample Fit (b)</th>
<th>Out Sample Fit (c)</th>
</tr>
</thead>
<tbody>
<tr>
<td>1 Standard Probit Model</td>
<td>$z_i = x_i^1 \beta^1 + x_i^2 \beta^2 + x_i^3 \beta^3 + x_i^4 \beta^4 + \varepsilon_i$</td>
<td>-237.681</td>
<td>0.177</td>
</tr>
<tr>
<td>2 Random Coefficients Model</td>
<td>$z_{ik} = x_i^1 \beta_k^1 + x_i^2 \beta_k^2 + x_i^3 \beta_k^3 + x_i^4 \beta_k^4 + \varepsilon_{ik}$</td>
<td>-203.809</td>
<td>0.158</td>
</tr>
<tr>
<td>3 Spatial model with geographic neighboring effect</td>
<td>$z_i = x_i^1 \beta^1 + x_i^2 \beta^2 + x_i^3 \beta^3 + x_i^4 \beta^4 + \theta_i + \varepsilon_i$</td>
<td>-146.452</td>
<td>0.136</td>
</tr>
<tr>
<td>4 Spatial model with geographic neighboring effect</td>
<td>$\theta_i = \rho \sum_{j=1}^{m} w_{ij} \theta_j + u_i$</td>
<td>-148.504</td>
<td>0.135</td>
</tr>
<tr>
<td>5 Spatial model with demographic neighboring effect</td>
<td>$z_i = x_i^1 \beta^1 + x_i^2 \beta^2 + x_i^3 \beta^3 + x_i^4 \beta^4 + \theta_i + \varepsilon_i$</td>
<td>-151.237</td>
<td>0.139</td>
</tr>
<tr>
<td>6 Spatial model with both geographic and demographic neighboring effect</td>
<td>$\theta_i = \rho \sum_{j=1}^{m} w_{ij} \theta_j + u_i$</td>
<td>-133.836</td>
<td>0.127</td>
</tr>
</tbody>
</table>

(a) $y_i = 1$ if a Japanese brand car is purchased (i indexes person and k indexes zip code)

- $x_i^1 = [1, \text{difference in price, difference in option}]$
- $x_i^2 = [\text{age, income, ethnic group, education}]$
- $x_i^3 = [\text{longitude, latitude}]$
- $x_i^4 = [\text{group1, group2, \ldots, group19}]$

Note that for other subgroups, estimation of these group level effects are not feasible because of the perfect or close to perfect correlation between the group dummy variable and $y$.

(b). In sample model fit is measured by the log marginal density calculated using the importance sampling method of Newton and Raftery (1994, P. 21).

(c). Our sample fit is measured by mean absolute deviation of estimated choice probability and actual choice.
### Table 5:
Posterior Mean of Coefficient Estimates

<table>
<thead>
<tr>
<th>Coefficient</th>
<th>Model 1</th>
<th>Model 2</th>
<th>Model 3</th>
<th>Model 4</th>
<th>Model 5</th>
<th>Model 6</th>
</tr>
</thead>
<tbody>
<tr>
<td>Intercept</td>
<td>-0.799</td>
<td>1.015</td>
<td>1.135</td>
<td>1.489</td>
<td>1.116</td>
<td>1.274</td>
</tr>
<tr>
<td>Price</td>
<td>-0.004</td>
<td>-0.003</td>
<td>-0.037</td>
<td>-0.039</td>
<td>-0.042</td>
<td>-0.038</td>
</tr>
<tr>
<td>Option</td>
<td>0.010</td>
<td>0.005</td>
<td>0.031</td>
<td>0.034</td>
<td>-0.028</td>
<td>-0.015</td>
</tr>
<tr>
<td>Age</td>
<td>-0.014</td>
<td>-0.014</td>
<td>-0.126</td>
<td>-0.119</td>
<td>-0.153</td>
<td>-0.080</td>
</tr>
<tr>
<td>Income</td>
<td>0.014</td>
<td>0.001</td>
<td>0.142</td>
<td>0.146</td>
<td>0.141</td>
<td>0.167</td>
</tr>
<tr>
<td>Education</td>
<td>-0.279</td>
<td>-0.032</td>
<td>-3.662</td>
<td>-4.003</td>
<td>-3.914</td>
<td>-3.425</td>
</tr>
<tr>
<td>Latitude</td>
<td>-2.901</td>
<td>-5.778</td>
<td>-6.397</td>
<td>-6.965</td>
<td>-6.231</td>
<td>-6.841</td>
</tr>
<tr>
<td>Group1</td>
<td>-1.069</td>
<td>-1.214</td>
<td>-0.897</td>
<td>-0.765</td>
<td>-0.682</td>
<td>-0.552</td>
</tr>
<tr>
<td>Group2</td>
<td>-1.591</td>
<td>-1.123</td>
<td>-1.159</td>
<td>-1.264</td>
<td>-1.378</td>
<td>-1.432</td>
</tr>
<tr>
<td>Group3</td>
<td>-1.142</td>
<td>-1.256</td>
<td>-0.569</td>
<td>-0.512</td>
<td>-0.165</td>
<td>-0.273</td>
</tr>
<tr>
<td>Group4</td>
<td>-1.207</td>
<td>-0.954</td>
<td>-3.159</td>
<td>-3.641</td>
<td>-2.331</td>
<td>-4.076</td>
</tr>
<tr>
<td>Group5</td>
<td>-0.499</td>
<td>-0.421</td>
<td>-2.315</td>
<td>-1.967</td>
<td>-1.578</td>
<td>2.523</td>
</tr>
<tr>
<td>Group6</td>
<td>-0.703</td>
<td>-0.946</td>
<td>-2.213</td>
<td>-2.786</td>
<td>-1.922</td>
<td>2.165</td>
</tr>
<tr>
<td>Group7</td>
<td>-0.604</td>
<td>-0.787</td>
<td>-1.887</td>
<td>-1.365</td>
<td>-2.134</td>
<td>2.632</td>
</tr>
<tr>
<td>Group8</td>
<td>-0.238</td>
<td>-0.113</td>
<td>1.459</td>
<td>1.781</td>
<td>0.967</td>
<td>0.957</td>
</tr>
<tr>
<td>Group9</td>
<td>-0.445</td>
<td>-0.335</td>
<td>2.646</td>
<td>2.893</td>
<td>1.384</td>
<td>2.484</td>
</tr>
<tr>
<td>Group10</td>
<td>-0.597</td>
<td>-0.667</td>
<td>0.132</td>
<td>0.197</td>
<td>0.268</td>
<td>0.169</td>
</tr>
<tr>
<td>Group11</td>
<td>-1.207</td>
<td>-1.016</td>
<td>-3.376</td>
<td>-3.132</td>
<td>-2.711</td>
<td>-2.954</td>
</tr>
<tr>
<td>Group12</td>
<td>-1.634</td>
<td>-1.147</td>
<td>-4.105</td>
<td>-3.198</td>
<td>-3.452</td>
<td>-4.103</td>
</tr>
<tr>
<td>Group13</td>
<td>-0.333</td>
<td>-0.178</td>
<td>2.139</td>
<td>2.164</td>
<td>2.551</td>
<td>2.872</td>
</tr>
<tr>
<td>Group14</td>
<td>-1.669</td>
<td>-1.124</td>
<td>2.778</td>
<td>2.115</td>
<td>2.160</td>
<td>2.949</td>
</tr>
<tr>
<td>Group15</td>
<td>-0.630</td>
<td>-0.421</td>
<td>1.986</td>
<td>1.468</td>
<td>1.235</td>
<td>1.836</td>
</tr>
<tr>
<td>Group16</td>
<td>-0.981</td>
<td>-0.563</td>
<td>-2.071</td>
<td>-2.569</td>
<td>-2.135</td>
<td>-3.088</td>
</tr>
<tr>
<td>Group17</td>
<td>-0.577</td>
<td>-0.289</td>
<td>1.468</td>
<td>1.923</td>
<td>1.497</td>
<td>2.313</td>
</tr>
<tr>
<td>Group18</td>
<td>-1.018</td>
<td>-0.956</td>
<td>0.912</td>
<td>0.813</td>
<td>0.569</td>
<td>0.783</td>
</tr>
<tr>
<td>Group19</td>
<td>-0.856</td>
<td>-0.433</td>
<td>1.876</td>
<td>1.579</td>
<td>1.692</td>
<td>0.964</td>
</tr>
<tr>
<td>$\sigma^2$</td>
<td>NA</td>
<td>NA</td>
<td>16.784</td>
<td>16.319</td>
<td>15.691</td>
<td>18.235</td>
</tr>
<tr>
<td>$\rho$</td>
<td>NA</td>
<td>NA</td>
<td>0.473</td>
<td>0.462</td>
<td>0.487</td>
<td>0.512</td>
</tr>
<tr>
<td>$\alpha$</td>
<td>NA</td>
<td>NA</td>
<td>NA</td>
<td>NA</td>
<td>NA</td>
<td>0.598</td>
</tr>
</tbody>
</table>

Note: Bolded figures indicate the 0 lies outside of the 95% highest posterior density interval of the estimate.
Figure 1.
Actual vs. Posterior Estimates of $\theta$ in the Numerical Simulation
Figure 2.
Histogram of Sample Size for Each Zip Code
Figure 3.
Estimated $\theta$ for Each Individual (666 people in total)

Note that the origin for longitude and latitude is –110 and 30. All the longitude and latitude information presented here is related to this origin specification.

Profiling the Two Groups

<table>
<thead>
<tr>
<th>Average</th>
<th>$\theta &gt; 0$</th>
<th>$\theta \leq 0$</th>
<th>P-value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Age</td>
<td>48.297</td>
<td>48.357</td>
<td>0.962</td>
</tr>
<tr>
<td>Income</td>
<td>69.835</td>
<td>71.934</td>
<td>0.418</td>
</tr>
<tr>
<td>Ethnic</td>
<td>0.116</td>
<td>0.094</td>
<td>0.275</td>
</tr>
<tr>
<td>Education</td>
<td>0.372</td>
<td>0.379</td>
<td>0.673</td>
</tr>
<tr>
<td>Latitude</td>
<td>4.152</td>
<td>4.158</td>
<td>0.502</td>
</tr>
<tr>
<td>Longitude</td>
<td>–8.281</td>
<td>-8.186</td>
<td>0.031</td>
</tr>
</tbody>
</table>