A Bayesian Structural Uncertainty Model to Target Loyalty and Conquesting Rebates to Consumers with Correlated Preferences

Bikram Karmakar
(University of Florida)

Ohjin Kwon
(Central Connecticut State University)

Gourab Mukherjee
(University of Southern California)

S. Siddarth
(University of Southern California)

Jorge M. Silva-Risso
(University of California, Riverside)

January 25, 2021

Bikram Karmakar is an Assistant Professor at the University of Florida; 102 Griffin-Floyd Hall, FL 32611, USA; email: bkarmakar@ufl.edu. Ohjin Kwon is an Associate Professor of Marketing at Central Connecticut State University. 1615 Stanley Street, New Britain, CT 06050; email: ohjin.kwon@ccsu.edu. Gourab Mukherjee is an Assistant Professor at the University of Southern California's Marshall School of Business; 3670 Trousdale Pkwy, Los Angeles, CA 90089-0809, USA; email: gourab@marshall.usc.edu. S. Siddarth is an Associate Professor of Marketing at the University of Southern California's Marshall School of Business; 701 Exposition Blvd, Los Angeles, CA 90089, USA; email: siddarth@marshall.usc.edu. Jorge M. Silva-Risso, Professor Emeritus, UC Riverside School of Business, Anderson Hall, 900 University Avenue, Riverside, CA 92521; email:jorge.silva-risso@ucr.edu.

A Bayesian Structural Uncertainty Model to Target Loyalty and Conquesting Rebates to Consumers with Correlated Preferences

ABSTRACT

We propose and estimate a spatial autoregressive multinomial probit model in which consumers' product preferences are correlated based upon their closeness to each other. Our proposed model uses a Bayesian structural uncertainty approach to combine multiple sources of such contiguity information and also incorporates consumer response heterogeneity. The model is applied to the unique problem of improving the efficacy of promotional programs that offer targeted conquesting and loyalty discounts to consumers, which is common in the auto industry but unstudied in the marketing literature.

Model calibration on automobile transaction data from the Los Angeles market confirms that previous purchases made by consumers are predictive of the future purchases of *other* consumers. Targeted discounts derived from the proposed model for conquesting and loyalty promotional programs substantially increase manufacturer profits. We demonstrate that the extant method of using a linear combination of the individual weight matrices provides an inferior fit and lower incremental profits than the proposed Bayesian structural uncertainty approach to information assimilation.

Keywords: *Structural uncertainty, Bayesian methods, Spatial choice model, Promotion Targeting*

JEL Classification: C11, C31, M31

1 Introduction

The lengthy inter-purchase times that characterize buying behavior in many durable goods markets (e.g. automobiles) imply that data on the past choices of individual consumers is non-existent or very sparse. Therefore, it is difficult to predict future choices based on consumers' past choice history and to design promotional targeting based on the predicted future choices. We propose a spatial autoregressive multinomial probit model of automobile choice wherein consumer preferences are correlated based on how similar or close they are to each other, and apply the estimated model to optimize customized rebates offered to consumers.

Levy (1959) claimed that products are symbols and people purchase them because of what they mean and communicate about the individual, not only for what they do. Product purchases may represent specific values that consumers seek (Reynolds and Jolly 1980, Sheth and Gross 1991), reflect the self-image of the consumer (Sirgy 1982), or convey some aspect of whom they want to be (Escalas and Bettman 2003), especially when these purchases are in publicly consumed product categories such as automobiles (Bourne 1957, Graeff, 1988). Escalas and Bettman (2005) find that the meaning of products used by reference groups may be appropriated to construct personal identity and self-concept (Belk 1988). Thus, if my member group of environmentally conscious consumers drives a Prius, then I may also do the same to reinforce my self-image.

This research suggests that purchases made by consumers may provide information about unobserved factors that affect future purchase decisions and, indeed, previous research has shown that past purchases of a brand are highly predictive of future brand purchases made by the *same* consumer. This phenomenon, termed structural state dependence (e.g. Erdem 1996, Seetharaman 2004), links a consumer's past purchases to his or her future choices of the same product. However, can my past purchases help to understand and predict the purchases of *other* consumers?

Our research answers this important question by proposing an estimable spatial discrete choice model, which leverages the similarity between consumers to infer their product preferences. Model parameters estimated on automobile transaction data reveal the predictive power of past purchases and other similarity measures. We demonstrate the practical value of the model by focusing on two types of consumer cash rebate programs, which are structured around the vehicle currently owned by a potential buyer. These programs are popular in the auto industry but hitherto unstudied in the marketing literature. The first, termed a loyalty rebate, rewards consumers who currently own the *same* vehicle model (e.g. a $500 rebate on a Camry for Camry owners), and the second, a so-called conquesting rebate, can only be availed by consumers who currently own a *different* vehicle model (e.g. a $500 rebate on a Camry for current Accord owners). We show how to use the model to optimize targeted discounts offered to consumers.

A vast literature in sociology and psychology suggests that homophily, the tendency for people to associate with similar others, also causes consumers who are close to one another in a social network to act in similar ways (Shalizi and Thomas, 2011). In marketing, Ma et al. 2015, use consumer purchases of call back ringtones to disentangle the effect of latent homophily from social influence and common exogenous factors. While our model also leverages the similarity between consumers to infer these correlated preferences, it does so by explicitly modeling the spatial correlation in the error term, *without* relying on the explicit ties between consumers that are the hallmark of homophily.

The potential predictive power of consumer similarity is easy to visualize in a situation in which the proximity between consumers stems from the physical distance between them. This geographic closeness, a proxy for many socio-demographic variables like income, education, wealth and property values, which are also related to consumer purchase behavior, has been the primary focus of previous spatial models in marketing (Yang and Allenby, 2003, Jank and Kannan, 2005) and economics (LeSage, 2003).

We contribute to the literature by expanding the scope of contiguity to incorporate the similarity in the products that consumers' have previously purchased and establishing the predictive power of doing so. The proposed model results in an estimable choice model in which consumer choice decisions are correlated *between* consumers, which is different from the *within*-consumer inertia effects documented by previous research in marketing.

Spatial models provide a natural way to model the correlation between different units of analysis based on how close they are in a similarity space, with most marketing applications focusing on physical proximity and using continuous rather than discrete outcomes. In their most basic form, these models use the spatial locations of the units or, more precisely, their proximity, to infer the correlation between them. Bronnenberg and Sismeiro (2002) and Bronnenberg and Mela (2004) model the correlation in brand shares across different markets. Bezawada et al. (2009) show how aisle placement of a brand in a particular store can affect the sales of another brand in a completely different category while Duan and Mela (2009) document the location-based correlation in housing prices.

In contrast, research on spatial correlation in preferences in a discrete choice setting is sparse. Jank and Kannan (2005) use experimental data to estimate a logit model of consumer choice of two product forms of a book - print or PDF - in which preferences and price sensitivities are spatially correlated across geographic regions. Yang and Allenby (2003) estimate a binary choice model in which consumer preferences for a vehicle's country-of-origin (Japanese/non-Japanese) are spatially correlated based on the distance and demographic similarity between consumers. The previously cited paper by Ma et al. (2015), models consumers' decision of whether or not to purchase a callback ringtone and which ringtone to acquire.

Our proposed model adds to this existing literature in both methodological and substantive ways. First, we propose and implement a theory-based innovative approach to combine multiple sources of consumer contiguity information based on the structural model uncertainty literature (Draper 1995, Kamakura et al. 1996). Specifically, we combine interpretable probabilities that emerge from each unique contiguity criterion thus coherently assimilating the associated predictive uncertainty, and validate it in our empirical application. Also, in contrast to the previously cited research in spatial models, the proposed spatial model allows for consumer heterogeneity in the marketing mix coefficients, which is an important requirement for any targeted or customized marketing campaign (Rossi et al. 1996).

Our substantive contribution is to add to the research on targeted promotional rebates in the automobile industry and on the role of trade-ins. Most previous trade-in research has studied how trading-in affects consumers' willingness to pay for a new car (Kwon et al. 2015). The downstream impact of the structure of trade-in contracts (Kim et al. 2011) and trade-in incentives (Miller et al. 2019) have also been studied. We add to this literature by developing a new similarity measure based on the features of traded-in vehicles and establishing its validity as a contiguity metric. Previous research in the promotion planning area (e.g., Silva-Risso and Ionova, 2008) uses a random coefficient logit model by relying on an a-priori geographic aggregation approach as the basis for improving the allocation of promotional incentives. Our model extends this approach by providing a more nuanced identification of a focal consumer's preference based on her proximity to other consumers, while accommodating response heterogeneity based on geographic aggregation. Further, we apply the model to improve the design of targeted discounts, termed loyalty and conquesting rebates, which are popular in the US auto industry but have not been previously studied in marketing.

We estimate the model parameters on a dataset of new car purchases in the Los Angeles market using a Hierarchical Bayes approach. The in-sample fit of the proposed model is superior to several other benchmark models, including the random coefficient probit model (RCP), spatial models that only include geographic closeness, or that directly combine contiguity matrices. The performance of the model in holdout samples is noteworthy; it yields a better fit than an RCP that has been calibrated directly on data for that market. We apply the model to improve a manufacturer's loyalty and conquesting programs and show that it yields higher profits than those from a state-of-the-art approach based on the RCP model (Rossi et al., 1996, Silva-Risso and Ionova 2008).

The rest of the paper proceeds as follows. In the next section, we provide details of the model. We then describe the dataset used in our empirical application and our Hierarchical Bayes estimation procedure, which is followed by a discussion of the estimation results from the calibration data and holdout sample performance. We then apply the parameters of the proposed model to the problem of improving a manufacturer promotion program in which targeted conquesting and loyalty discounts are offered to consumers. We conclude with a summary of our findings and some directions for future research.

2 Model

We begin with a very general specification of the probit choice model, which forms the basis of a number of non-spatial and spatial models that are estimated in the empirical part of the paper.

Figure 1 provides a schematic overview of the basic components and the models while Table 1 summarizes their key differences.

==

Insert Table 1 and Figure 1 about here

==

The outcome variable that we focus on is the vehicle model $j = 1, 2, ..., J$ chosen by consumer $i = 1, 2, \dots, n$, conditional on purchase in the product category. This choice outcome, y_i, is modeled using an additive random utility framework, yielding

(1) $$y_i = j^* \; if \; U_{ij^*} > U_{ij} \text{ for all } j \neq j^*,$$

where U_{ij} are the endogenous latent utilities for each consumer-vehicle combination. A heteroskedastic regression model with unknown variance vector, σ_j, is used to specify U_{ij} as

(2) $$U_{ij} = \mu_{ij} + X_{ij}\beta_i + \sigma_j\varepsilon_{ij} \;\; for \;\; i = 1, \dots, n \text{ and } j = 1, \dots, J \; ,$$

where X_{ij} is a matrix of alternative- and consumer-specific variables, μ_{ij} and β_i are consumer-specific preference and response parameters, respectively, and ε_{ij} are i.i.d. normal errors. Fixing J as the reference vehicle for model identification purposes implies that equations (1)-(2) yield a generalized multinomial probit model. Because only one outcome y_i is observed for consumer i, this model is over-parameterized and additional structural assumptions must be imposed to estimate the parameters efficiently. Models 1 and 2 in Table 1, the basic probit and the RCP model, respectively. Both are specified and estimated in the standard way though it is important to note that for the RCP heterogeneity is incorporated at the ZIP code level. We now turn to more details about the spatial models.

To isolate the benefits provided by modeling spatially correlated preferences and to clarify model exposition, we start by describing Models 3, 4 and 5 in Table 1 that incorporate these preferences but *do not* incorporate response heterogeneity.

Let W be a $n \times n$ weight matrix that is based on a well-defined contiguity characteristic (e.g., spatial distance), whose individual elements are $w_{ii} = 0, w_{ik} \geq 0$ and $\sum_{k=1}^{n} w_{ik} = 1$. The last constraint ensures that the contiguities are always row-normalized so w_{ik} is the relative contiguity of individual k with individual i. To incorporate spatially correlated preferences (SCP) (but not response heterogeneity) the following structural constraints are imposed on the parameters in equations (1)-(2):

(3) $$\beta_i = \beta \text{ and } \mu_{ij} = \alpha_j + \theta_{ij}.$$

(4) $$\theta_{ij} = \rho \sum_{k=1}^{n} w_{ik} \theta_{kj} + \psi_j \xi_{ij}.$$

This implies that consumer utility, U_{ij} in equation (2), consists of a deterministic part $\alpha_j + X_{ij}\beta$, with α_j common to all consumers, a stochastic component, θ_{ij}, which represents the correlated component of consumer i's preference for alternative j, and the probit error term, ε_{ij}.

Based on the mixed regressive, spatial autoregressive model of Anselin (1988), we impose a second-level hierarchical structure on the θ_{ij} in (4), and specify the preference vector of each consumer, θ_{ij}, as a weighted average of the preferences of all other individuals with the similarity between i^{th} and k^{th} consumers serving as the weight. Thus, for two consumers, k and k', the former being more similar to i than the latter (i.e., $w_{ik'} < w_{ik}$), equation (4) ensures that consumer i's preferences will correlate more highly with k than with k'.

The spatial autoregressive parameter, ρ, captures the average influence of neighboring consumers on the preferences of the focal consumer. The residuals ξ_{ij} in (4) have a standard normal distribution and heterogeneity in the variances is admitted through the unknown standard deviation ψ_js. The key difference between the spatial autoregressive model in (4) and the probit model of Jank and Kannan (2005) (henceforth referred as JK), which includes two alternatives and an outside good with spatially correlated utilities, is that our model includes a spatial multiplier, ρ, which allows for global spillovers of the spatial structure across all the covariates, while JK's model does not.

All of the proposed spatial choice models rely on the spatial weight matrix W as a critical building block. For a set of n consumers, each element w_{ik} of the $n \times n$ matrix, W represents the proximity between a "row" consumer (i) and a "column" consumer (k). In our empirical application, W is based on either of two natural choices for proximity between consumers: (i) geographical location, which yields matrix W^G consisting of individual entries, w_{ik}^G and (ii) how similar their previously purchased vehicles are to each other, which yields W^V, with individual elements w_{ik}^V.[1] Models 3 and 4 use W^G and W^V, respectively, to incorporate spatially correlated preferences.

The availability of multiple contiguity matrices naturally raises the question of how best to combine information from them. We start by recognizing that each contiguity matrix is separately informative about preference correlation and the vector of choice probabilities that emerge from each W represents the final effect of the closeness information. Also, because our targeting application requires that we leverage all of the information available in the matrices maximally, it is critically important that the interpretable probabilities tied to each unique contiguity criterion be combined in a structured manner that ensures that the contiguity criteria more predictive of each consumer's preferences receives a proportionately greater weight.

---

[1] We postpone a detailed description of how the individual elements of each matrix are constructed to the empirical application section.

Noting these requirements, we propose an innovative, structural, approach to combine the information in the two matrices using Bayesian model averaging to mix the individual probabilities associated with W^G and W^V. Not only does Bayesian model averaging provide a rigorous mechanism for incorporating the structural uncertainty associated with the different models (Carlin and Chib, 1995; Diebolt and Roberts, 1994; Wedel and Sabro, 1994; Steel 2020), it also allows the information to be combined non-linearly and effectively incorporates heterogeneity in the contiguity criterion. A detailed description of the method follows.

Consider L probit models M_ls with spatially correlated preferences based on (1)-(4), each associated with a corresponding weight matrix $W^{(l)}$. Each model M_l defines a set of probabilities $\tau_{ij}(l)$s of a customer i choosing product j. For each consumer i, we would like to combine the individual's $\tau_{ij}(l)$s from different models. Following Hoeting et al. (1999), it follows that $\tau_{ij}(l) = P(M_l, Y)P(M_l|\ Y)$ where Y is the observed choice vector and $\tilde{Y} = (\tilde{y}_1, \dots, \tilde{y}_n)$ is the future choice vector based on model (1)-(4). If each model is assumed to be a-priori equally likely we have $P(\mathcal{M}_l|Y) = P(Y|\mathcal{M}_l)/\{\sum_k P(Y|\mathcal{M}_k)\}$ where $P(Y|\mathcal{M}_l)$ is the integrated Bayes risk, written in terms of the parameters in (1)-(4) as:

(5) $$P(Y|\mathcal{M}_l) = \int P(Y|\alpha, \beta, \theta, \sigma)\, \pi(\alpha)\, \pi(\beta)\, \pi\left(\theta \middle| W^{(l)}, \psi\right) \pi(\psi)\, \pi(\sigma)\; d\alpha\; d\beta\; d\theta\; d\psi\ ,$$

and all subscripts are omitted to improve clarity. In the above $P(\alpha, \beta, \theta, \sigma)$ is the likelihood based on the model (1)-(3) and $\pi\left(W^{(l)}, \psi\right)$ is the prior from (4) based on the weight matrix $W^{(l)}$. These $P(M_l)$ values are combined with a set of mixing parameters for the L models which are non-negative and sum to unity. In our empirical study with two weight matrices, there is only one free mixing parameter. We use a beta prior over this mixing parameter. Independent priors are used for the intercepts α, response parameters β as well as the spreads σ and ψ. The Appendix provides a detailed MCMC algorithm for implementing this approach.

We note that the extant method (e.g., Yang and Allenby, 2003) weights the component contiguity matrices differently to produce a single weight matrix: In this method, $W^C = \sum_{l=1}^{L} \lambda_l W^{(l)}$ where, $\lambda_l \geq 0$, $\sum_{l=1}^{L} \lambda_l = 1$. A weakness of this approach is that if one of the contiguity matrices, say W^V, is more predictive of the preference correlation, then the model fails to recognize that a larger fraction of the population should be modeled using W^V. In contrast, because the Bayesian model averaging approach mixes probabilities, as opposed to contiguity, it ensures that more weight is given to W^V. The empirical application demonstrates that the Bayesian model averaging is superior on predictive performance and in capturing structural uncertainty.

A vast literature in marketing, and the automobile industry, in particular, has shown the importance of accounting for heterogeneity in modeling consumer demand particularly for designing targeted promotional programs (Rossi et al. 1998, Silva-Risso and Ionova 2008). In models estimated for CPG products using scanner panel data, the household is the typical unit of analysis because multiple observations are available at this disaggregate level. In durable goods markets such as cars, with inter-purchase times of about five to seven years, however, some form of geographical aggregation is typical (Bucklin et al. 2008). While the preference coefficients in Model 6 already account for such heterogeneity, the response coefficients do not. Therefore, we propose Model 7, which allows for ZIP code level heterogeneity in the response parameters so that $\beta_i = \beta_k \; if \; zc_i = zc_k$ for any $1 \leq i, k \leq n$ and for all $j = 1, \ldots, n$. In so doing, we not only add to the spatial model literature but also the literature on promotional planning. Figure 1 shows how the various components of the model are combined into the final Model.

3 Data and Estimation

3.1 Data

The Power Information Network (PIN), a division of J. D. Power and Associates, collects new vehicle transaction data from a large number of dealers electronically. The data used in this study come from a transaction history of new car purchases in the midsize sedan category made by consumers in the Los Angeles Designated Market Area (DMA) during the first six months of 2007.[2]

Since the four top-selling models accounted for about 93% of all transactions in this category, our analysis is restricted to predicting consumer choice among these four models. Further, because we are interested in examining the impact of previously purchased vehicles, we restrict our sample to only those transactions in which a consumer purchasing a new car also traded-in another vehicle. Although consumers in our sample purchased only one of the four shortlisted midsize sedans, they traded-in 295 different vehicle models representing 22 manufacturers.

The final data set consisted of a total of 1342 new car transactions over a six-month period: 821 transactions from the first four months are used to calibrate model parameters and the remaining 521 transactions are held out for model validation in the Los Angeles market.

The data includes several details of each transaction including the price each individual consumer paid for the vehicle, the Annual Percentage Rate (APR) for finance and lease contracts, the monthly payment amount, manufacturer rebate (if any) and the residual value of the vehicle if it was leased. The data also contain the geo-coded location (i.e., precise latitude and longitude coordinates) of the consumer's residence as well as detailed attribute information for traded-in

[2] All statistics are based on the observations in our data set.

vehicles. Table 2 reports descriptive statistics for our sample including vehicle market shares expressed as a fraction of the sales of the included models.

==

Insert Table 2 about here

==

The variables included in the choice models are the log of price and a last-make dummy variable (1 if make of the new car (e.g. Toyota) is the same as that of the traded-in car), which captures the inertia or state dependence in vehicle choice, and the distance between a consumer's home address and the closest dealer of the car make. We standardized vehicle transaction price to construct the baseline price net of vehicle options with a hedonic regression (e.g. Zettelmeyer et al. 2006). We then constructed adjusted vehicle prices by subtracting manufacturer rebate, the dollar amount of APR promotion, and trade-in over- or under- allowance. We calculated the dollar amount of an APR promotion using a 5% base APR level. Specifically, for transactions with APR's less than 5%, the dollar amount corresponding to the APR subvention was treated as an APR promotion. APR's greater than 5% were considered non-promotional and the dollar amount of the promotion was set to zero. The data also includes a field for Trade-in over-[under-] allowance, which represents the difference between the price the dealer pays to the consumer for the trade-in car and its wholesale value. Paid prices are adjusted by the over-[under-] allowance to control for the possibility that dealers may pay consumers a higher[lower] price for the traded-in vehicle and then charge them a lower[higher] price for the new car (Scott-Morton et al. 2001).

Because retailers' pricing decisions can be made based on market conditions or vehicle characteristics that are unobservable by researchers, prices and the probit error term may be correlated. The potential endogeneity of prices may bias the estimate of the price coefficient

(Villas-Boas and Winer 1999). We address this issue using the control function approach of Petrin and Train (2010), i.e., by first regressing transaction prices for each vehicle model on instruments and using the residuals from these equations for each vehicle model as additional explanatory variables in the choice model. We use the wholesale prices for each vehicle model (i.e. price that a dealer pays for a car to a manufacturer) as instruments because they are highly correlated with retail price but unlikely to be correlated with unobservables that affect retailer pricing decisions. This procedure produces three price residual variables corresponding to Accord, Altima and Camry with Passat being a baseline alternative.

3.2 Similarity Matrices

We now describe how the two W matrices in our study were operationalized.

3.2.1 Geographic Location

Each element of W^G is calculated using a two-step approach. Let d_{ik} represent the Euclidian geographic distance between the residential locations of consumers i and k. In the first step, a raw contiguity measure between these consumers is calculated as: $\widetilde{w}_{ik} = \exp(\,1/d_{ik})$ if $i \neq k$ and $w_{ii} = 0$, which ensures that contiguity increases as the distance between the residential locations decreases (Bezawada et al. 2009, Yang and Allenby 2003). In the second step, the raw $\widetilde{w}_{ik}$ values are row-normalized, so values in each row sum to unity as in Anselin (1988) and LeSage (2000). The final elements of the matrix are given by $w_{ik}^G = \exp(1/d_{ik})/\{\sum_{l \neq i} \exp(1/d_{il})\}$ for $i \neq k$ and 0 for $i = k$.

3.2.2 Previous Vehicle

We use a Lancasterian perspective to represent the consumer's previously-owned vehicle as a vector of attributes. Specifically, each trade-in vehicle is described by the following five characteristics: manufacturer (e.g. GM, Honda.), nameplate (e.g. Chevrolet, Acura.), model (e.g. Taurus, Accord.), manufacturer continent of origin (e.g. American, Asian) and the number of Cylinders (4, 6, 8), and the raw similarity measure is specified as $\exp(c_{ik})$, where c_{ik} is the number of characteristics common to the vehicles traded in by consumers i and k. For example, if consumers i and k traded in a Chevrolet Malibu and a Buick Regal, respectively, then, because the two vehicles share the same continent of origin and manufacturer, $c_{ik} = 2$. This operationalization implicitly assumes that all vehicle characteristics are equally important, which reduces the number of parameters required to estimate the model (Ho and Chong 2003). As before, the raw elements of the matrix are normalized so that the sum of the elements in each row is one, and the diagonal elements of W are zero. Thus, the final elements of the vehicle similarity matrix are given by $w_{ik}^{V} = \exp(c_{ik})/\{\sum_{l \neq i} \exp(c_{ik})\}$ for $i \neq k$ and 0 for $i = k$, which is consistent with the discussion in Bavaud (1998). The exponential function ensures positive values in the similarity matrix.

3.3 Estimation

We used Markov Chain Monte Carlo (MCMC) methods to estimate the parameters of each model. Specifically, we employed Gibbs sampling steps to make draws from the full conditional distributions of each parameter, except for ρ, which required the use of a Metropolis-Hastings step because its posterior distribution did not have a closed-form solution. We ran the sampler for 100,000 iterations, thinning it by retaining every 10th draw, and assessed convergence by monitoring the time-series plots of the parameter draws. We discarded the first 9,000 retained draws as "burn-in" and used the last 1,000 draws to make inferences about the posterior distribution of the parameters. We checked the stability of our estimates by comparing them against two other samples in which every draw or every fifth draw after burn-in, respectively, was retained for posterior inference. For Model 5, the two weight matrices corresponding to vehicle and geographic respectively are combined using a beta prior on the mixing weight. Details of the estimation procedure, including the conditional distributions of each of the parameters as well as the sampling techniques involved in structural uncertainty based aggregation of Models 6 and 7 are presented in the Appendix.

Our algorithm was coded using statistical software R. On a 1.7GHz quad-core Intel i7 processor with 32GB RAM, 100,000 MCMC iterations for Models 3, 4 and 5 took about 30 hours and about 50 hours for Models 6 and 7. For each model, we verified convergences using various statistics and trace plots that are available the R packages coda (Plummer et al. 2019) and mcmcplot (Curtis et al. 2015).

4 Results

4.1 Los Angeles Data

To evaluate the performance of the proposed model, we used the calibration sample to estimate the parameters for the seven different models. Two non-spatial models serve as the benchmark, the standard probit model (Model 1), and the random coefficient probit model (Model 2), which allows for the usual preference and response heterogeneity using the a priori aggregation approach, with parameters aggregated to the three-digit ZIP code level.[3]

Five different spatial models are fit to the data. Models 3 and 4 represent two different versions of the Preference Correlation model in which the preference coefficients (intercept terms in the utility function) of consumers are spatially correlated based on geographic contiguity (Model 3) and vehicle similarity (Model 4), respectively. Models 5 and 6 allow for both spatial contiguity matrices to affect the spatial correlation in preference coefficients. In Model 5 the final W matrix is a convex combination of the individual, geographic- and vehicle-similarity based contiguity matrices, a la Yang and Allenby (2003). Model 6, based on the structural uncertainty approach, mixes the individual probabilities derived from each W matrix. Finally, Model 7 incorporates the spatial correlation and mixing of probabilities in Model 6, but also allows for heterogeneity in the response coefficients by aggregating observations for each three-digit ZIP code. To make predictions from the spatial models on holdout samples in geographical markets, the preference and marketing mix parameters for consumers in each target market must be imputed from the parameters estimated in the calibration sample. Next, we describe how this is done.

4.2 Imputation Procedure for Holdout Samples

The following two-step imputation procedure is used to impute the preference parameters. In the first step, an augmented weight matrix (AW) is created to include consumers from both calibration

[3] The data set contains 24 three-digit ZIP codes, yielding an average of about 48 observations for each geographic unit. We prefer this level of aggregation because using five-digit ZIP codes instead, results in some ZIP codes having only very few observations.

and holdout samples. Thus AW has dimension 1342×1342[4], with individual entries representing the vehicle similarity between row and column consumers. In the second step, the preference parameters for the holdout consumers are obtained by averaging the corresponding values for consumers in the calibration sample, with the corresponding elements of the AW matrix serving as weights. In other words, we use the estimated parameters from the calibration sample but weight them by the (normalized) similarity between the calibration and hold out consumers. Thus, for a holdout consumer h, we set:

(6) $$\hat{\theta}_{hj} = \sum_c w_{hc}\hat{\theta}_{cj}\,,\ \hat{\mu}_{hj} = \sum_c w_{hc}\hat{\mu}_{cj} \ \text{ for } j = 1, \ldots, J,$$

where c varies overall calibration consumers and w_{hc} are the similarity weights between consumer c in the calibration sample and consumer h. The marketing mix parameters for the holdout sample are directly drawn from the estimated ZIP code specific posterior distribution.

4.2 Model Performance Comparison

Following previous research (Allenby et al. 1998) model performance is evaluated using two measures: log marginal density (LMD) of Newton and Raftery (1994) calculated after eliminating very low likelihood regions and the mean absolute deviation (MAD). These fit statistics are reported in Table 3 for the Los Angeles data.

==

Insert Table 3 about here

==

[4] 821 consumers in the calibration sample and 511 consumers in the holdout sample.

The results reveal that, as expected, Model 2 fits the data better than Model 1, establishing that consumer heterogeneity plays a significant role in new car purchases. In the models accounting for spatial correlation in preference, Model 4 provides a better fit than Model 3, which implies that previous vehicle similarity is a more important driver of the spatial correlation in consumer preference than is geographic closeness. Interestingly, the direct mixing of W matrices, as in Yang and Allenby (2003), yields an *inferior* fit in the calibration sample (LMD –490.52 and MAD 0.325) than that obtained from only using the vehicle contiguity matrix (LMD –482.38 and MAD 0. 317).

Holdout sample fit is consistent with these results (MAD of 0.425 versus 0.423). The W matrix mixing parameter is estimated to be 0.514 in Model 5, which, while weighting the previous vehicle similarity matrix heavily, still deteriorates model fit. This indicates that the existing reduced-form method of combining contiguity matrices can fail to capture similarity information of consumers. In contrast, the structural uncertainty based choice probability mixing performs the best among the pure spatially correlated preference models: LMD –462.43 and MAD 0.315 in the calibration sample and MAD of 0.421 in the holdout sample. Finally, Model 7 yields the best fit in all samples, with a noticeably significant improvement in holdout MAD to 0.373. As such Model 7 produces 11.4% reduction in out-of-sample error from its nearest competitor, underscoring the value of incorporating spatial correlation in both preference and response heterogeneity. Therefore, the rest of our discussion is based on the parameter estimates of Model 7.

Mathematically, the main difference between Models 5 and 6 is that while the latter combines the two spatially autocorrelated models based on geographic and vehicle weights using their respective choice probabilities, the former directly combines the weight matrices. It is often difficult to construct an ideal linear combination of the weight matrices preserving the important local features of the individual weight matrices. We observe this in Table 3. In Table 3 we found that Model 5 performed slightly worse than Model 4 that uses only vehicle characteristics. In Model 4, λ_v and $\lambda_g = 1 - \lambda_v$ were set to 1 and 0 respectively whereas the optimal calibrated values of λ_v and λ_g in Model 5 were 0.54 and 0.46 respectively. We used a symmetric beta prior on λ_v in Model 5 and it appears that the Bayes risk based on the combination of weight matrices was not very sensitive to the values of the λ_v parameter which subsequently led to imprecise estimation.

On the other hand, Bayesian model averaging in Model 6 improves both estimation and predictive accuracy by using 0.95 and 0.05 weights to combine probabilities from vehicle and geographic characteristics based spatial autocorrelation models. The estimated λ_v weight in Model 6 reconfirms that vehicle characteristics are more useful than geographic contiguity as seen in model 3 and 4. But Model 6's significantly better performance over Model 4 also illustrates that in consumer subsets where the geographically weighted model works better than the vehicle model, its impact is not lost but properly reflected in their probabilistically weighted combination.

==

Insert Table 4 about here

==

Table 4 reports the coefficients from Model 7. The coefficient for the Last Make dummy is positive and highly significant, showing that consumers tended to purchase the same make vehicle as he or she had chosen in the previous purchase occasion. The estimate of the average price coefficients across all ZIP codes is – 1.58 and its high posterior variance show that price-sensitivity varies greatly across ZIP codes. At a 10% significance level, we see that the average price coefficient is significantly negative; but because of heterogeneity across ZIP codes, the average price coefficient is not significant at 5% level.

Spatial spillover autoregressive parameters for both geographic proximity and vehicular preference similarity are significant, which indicates that consumers' preferences for vehicle models are positively correlated with preferences of both geographically neighboring consumers and those who had chosen similar vehicles in a previous purchase occasion.

5 Determining Optimal Conquesting and Loyalty Rebates

In this section, we use the parameter estimates from the proposed model to find the optimal loyalty or conquesting rebate amount to be offered to each three-digit ZIP code. For comparison purposes, we use the parameters from Model 2 in two ways: (1) to derive the optimal rebate for each ZIP code not based on previous vehicle ownership (a la Silva-Risso and Ionova 1996), and (2) the equivalent optimal loyalty or conquesting rebate for each ZIP code.

Taking the position of a Toyota Camry marketing manager, we used the following procedure to derive the optimal rebate for each ZIP code from our model. The basic building block of this analysis is the predicted contribution $C(R_k, Z, t)$ from a rebate of type t (loyalty or conquesting), amount $\$R_k$, in ZIP code Z. The expected contribution of this rebate on the set of all consumers in ZIP code Z, who get this rebate, $C(Z, t)$, is calculated as:

(7) $1/L \sum_{l=1}^{L} \sum_{c \in C(Z,t)} Prob_l(\text{customer } c \text{ bought Camry} | Price_c, R_k) \times Margin(Price_c, R_k)$,

using the L retained draws from the posterior distribution of model parameters for the concerned ZIP code. The value k of the rebates R_k were varied in a grid starting from \$0 to the highest observable value, \$3200, in steps of \$100. Specifically, the probability that consumer c in ZIP code Z chooses the Camry when it offers a consumer cash rebate was based on the posterior parameter estimates from the l[th] retained draw, keeping the set of other covariates the same, but decreasing the price by the face value of the rebate. Because data on manufacturer margin is not readily available, based on our discussions with industry experts we assume it to be 25% of the selling price. The optimal rebate when a single rebate level is applied to all consumers in ZIP code Z is given by $R_Z^* = C(R_k, Z, all)$. The total profit from the optimal single rebate strategy is determined by summing these maximum contributions over all the ZIP codes, i.e.,

(8) $$C_{Total}^{S} = \sum_z C(\, R_Z^*, \mathrm{Z}, \mathrm{all}).$$

Now, consider rebate types that are based on the vehicle that a consumer currently owns. The optimal loyalty and conquesting rebate levels for ZIP code Z are obtained as

(9) $$R_{Z,L}^* = \underset{k}{\operatorname{argmax}}\, C(R_k, Z, \text{Camry owners}) \text{ and } R_{Z,C}^* = \underset{k}{\operatorname{argmax}}\, C(R_k, Z, \text{non} - \text{Camry owners}).$$

Unlike the single rebate strategy, the optimal face values, $R_{Z,L}^*$ and $R_{Z,C}^*$, can be different for a ZIP code Z and total profits for the optimal strategy are determined by summing the maximum contributions due to these loyalty and conquesting rebates over all ZIP codes.

(10) $$C_{Total}^{LC} = \sum_Z C\big(R_{Z,L}^*, \mathrm{Z}, \text{Camry owners}\big) + \; C\big(R_{Z,C}^*, \mathrm{Z}, \text{non} - \text{Camry owners}\big).$$

The profit per consumer from the optimal single rebate per ZIP code, not accounting for current vehicle ownership is $2894[5]. Figure 2(a) and 2(b), plot the histograms of the optimal loyalty and conquesting rebates, respectively, derived from Model 2, which yield a profit per consumer of $2952, an increase of about 2 percent over the base case. This difference shows that targeting rebates based on current vehicle ownership has some upside profit potential relative to a strategy that ignores this information. The impact is particularly striking for the strategy derived from Model 7, which yields an average profit of $3553 per ZIP code, or about 23% more than the

[5] For brevity the histogram of rebates associated with this strategy is not included.

baseline. In contrast, the optimal loyalty and conquesting rebate derived from Model 5 yields a much lower profit per consumer of $3063, which is about the same as those from Model 2.

The comparison of the histograms of rebates in Figure 2 reveals that the recommended targeting strategies from Model 2 (panels a and b) are very different from those from Model 7 (panels c and d). Specifically, the non-zero loyalty rebates from Model 2 (panel a) are more numerous and have a higher face value than those from Model 7 (panel c). Also, the conquesting rebates from Model 2 (panel b) have a noticeably smaller zero group and many face values that lie above $2500, while those from Model 7 (panel d) have many ZIP codes with no rebate and offered rebates that are of lower face values, the vast majority below $1500 and none above $2000.

==

Insert Figure 2 about here

==

6 Conclusion

We develop and estimate a new Bayesian spatial choice model that permits preference parameters to be spatially correlated among consumers and also incorporates heterogeneous response parameters. We apply the model to transaction data from the Los Angeles automobile market and find that the proposed model fits the data better than the RCP model and several benchmark models. While previous applications of spatial models in marketing have highlighted the role of geographic contiguity in demand prediction, we show that the similarity in consumers' previously purchased vehicles is even more important. Specifically, in our empirical application, we found that the estimated mixing parameter gave 95% of the weight to vehicle similarity and only 5% to geographic contiguity. While previous research has documented that past purchases are highly predictive of future choices made by a consumer, our research shows that a spatial model can help expand this predictive power beyond *within-consumer* inertia to *other* consumer's choices.

We make a methodological contribution to the spatial literature by proposing a new approach to assimilate predictive uncertainties associated with different kinds of weight matrices in a coherent and structured manner, which outperforms the existing approach. We also expand on the existing promotional targeting literature by studying discounts based on previous purchases, specifically loyalty- and conquesting rebates, which have been understudied in the literature. We show how managers can use the model to come up with better rebate programs, how these types of rebates are more effective than traditional price discounts that ignore previous purchases and, most importantly, that loyalty and conquesting rebates derived from the proposed model yield substantially higher profits.

There are several limitations to this research, which also present opportunities for future research in this area. First, in our work contiguity is only based on geographic distance and previous vehicle information, which could be expanded to include other socio-demographic or purchase characteristics. For example, consumer's income, ethnicity, age or previous search history or dealership visits may form a viable basis of W matrices that explain variation in intrinsic preferences. Similarities from the previous vehicle-based W matrix can be expanded to include more attributes, for example, engine size, gas mileage and so on. Second, our model only allows for spatial correlation in the preference parameters, so a promising avenue for future research is to develop models that can also incorporate spatial correlation in the response parameters. Third, the modeling approach could be extended to accommodate spatial correlation in the features of the alternatives themselves, similar to the spatial demand model Duan and Mela (2009) in which geographic distance between alternatives (outlet location) serves to identify consumer preferences.

Finally, while our empirical application demonstrates the value of modeling spatial correlation applying it to larger datasets would yield practical benefits to manufacturers with large product lines. Thus, scaling up the algorithms to do this more efficiently also represents a promising area of future research.

References

Allenby, G.M., Arora, N., Ginter, J.L. (1998). "On the Heterogeneity of Demand," *Journal of Marketing Research*, 35(3), 384-389.

Anselin, L. (1988). *Spatial Econometrics: Methods and Models.* Dorddrecht: Kluwer Academic Publishers.

Anselin, L. (2002). "Under the hood issues in the specification and interpretation of spatial regression models," *Agricultural economics*, *27*(3), pp.247-267.

Bavaud, F. (1998), "Models for Spatial Weights: A Systematic Look," *Geographical Analysis*, 30. 153-171.

Belk, R.W. (1988). "Possessions and the Extended Self," *Journal of Consumer Research*, 15, 139-168.

Bezawada R., Kannan, P.K., Balachander, S., Shankar, S. (2009). "Cross-Category Effects of Aisle and Display Placements: A Spatial Modeling Approach and Insights," *Journal of Marketing,* 73(3), 99-117.

Bourne, F.S. (1957). *Group Influence in Marketing and Public Relations: Some Applications of Behavioral Research.* UNESCO, Basel, Switzerland.

Bronnenberg, B.J., & Mela, C.F. (2004). "Market Roll-Out and Retailer Adoption for New Brands," *Marketing Science*, 23(4), 500-518.

Bronnenberg, B.J., & Sismeiro, C. (2002). "Using Multimarket Data to Predict Brand Performance in Markets for which No or Poor Data Exist," *Journal of Marketing Research*, 39(1), 1-17.

Bucklin, R.E., Siddarth, S., Silva-Risso, J.M. (2008). "Distribution Intensity and New Car Choice," *Journal of Marketing Research*, 45(5), 473-486.

Carlin, B.P., & Chib, S. (1995). “Bayesian model choice via Markov chain Monte Carlo methods,” *Journal of the Royal Statistical Society: Series B (Methodological)*, *57*(3), pp.473-484.

Curtis, S.M., Goldin, I., Evangelou, E. (2015). mcmcplots: Create plots from MCMC output. *R package version 0.4*, *2*.

Diebolt, J., & Robert, C.P. (1994). “Estimation of finite mixture distributions through Bayesian sampling,” *Journal of the Royal Statistical Society: Series B (Methodological)*, *56*(2), pp.363-375.

Draper, D. (1995). “Assessment and propagation of model uncertainty,” *Journal of the Royal Statistical Society: Series B (Methodological)*, *57*(1), pp.45-70.

Duan, J.A., & Mela, C.F. (2009). “The Role of Spatial Demand on Outlet Location,” *Journal of Marketing Research*, 46(2), 260-278.

Erdem, T. (1996). “A Dynamic Analysis of Market Structure Based on Panel Data,” *Marketing Science*, 15(4), 359-378.

Escalas, J.E., & Bettman, J.R. (2003). “You Are What They Eat: The Influence of Reference Groups on Consumers' Connections to Brands,” *Journal of Consumer Psychology*, 13(3), 339-348.

Escalas, J.E., & Bettman, J.R. (2005). “Self-Construal, Reference Groups, and Brand Meaning,” *Journal of Consumer Research*, 32(3), 378-389.

Graeff, T.R. (1988). “Consumption Situations and the Effects of brand image Influence of Reference Groups on Consumers’ Connections to Brands,” *Journal of Consumer Research*, 13(3), 339-348.

Ho, T., & Chong, J. (2003). “A Parsimonious Model of Stockkeeping-Unit Choice,” *Journal of Marketing Research*, 40(3), 351-365.

Hoeting, J.A., Madigan, D., Raftery, A.E., Volinsky, C.T. (1999). “Bayesian model averaging: a tutorial,” *Statistical Science*, pp.382-401.

Jank, W., & Kannan, P.K. (2005). “Understanding Geographical Markets of Online Firms Using Spatial Models of Customer Choice,” *Marketing Science*, 24(4), 623-634.

Kamakura, W.A., Kim, B.D., Lee, J. (1996). “Modeling preference and structural heterogeneity in consumer choice,” *Marketing Science*, *15*(2), pp.152-172.

Kim J, Rao R.S., Kim K, Rao A.R. (2011). “More or Less: A Model and Empirical Evidence on Preferences for Under- and Overpayment in Trade-In Transactions,” *Journal of Marketing Research*, 48(1):157-171.

Kwon, O., Dukes, A.J., Siddarth, S., Silva-Risso, J.M. (2015). “The informational role of product trade-ins for pricing durable goods,” *The Journal of Industrial Economics*, 63(4), 736-762.

LeSage, J.P. (2000). Bayesian Estimation of Limited Dependent Variable Spatial Autoregressive Models. *Geographical Analysis*, 32(1), 19-35.

LeSage, J.P. (2003). *A Family of Geographically Weighted Regression Models. Advances in Spatial Econometrics: Methodology, Tools and Applications*, Springer, New York.

Levy, S.J. (1959). “Symbols for Sale,” *Harvard Business Review*, 37 (July–August), 117–24.

Ma, L., Krishnan, R., Montgomery, A.L. (2015). “Latent Homophily or Social Influence? An Empirical Analysis of Purchase within a Social Network,” *Management Science*, 61(2), 454-473.

Manchanda, P., Ansari, A., Gupta, S. (1996). “The Shopping Basket: A Model for Multicategory Purchase Incidence,” *Marketing Science*, 18(2), 95-114.

Chadwick, M., Wiles, M., Park, S. (2019), “Trading On Up: An Examination of Factors Influencing the Degree of Upgrade: Evidence from Cash for Clunkers,” *Journal of Marketing*, 83 (1), 151-172.

Mittal, V., Kamakura, W.A., Govind, R. (2004). “Geographic Patterns in Customer Service and Satisfaction: An Empirical Investigation,”*Journal of Marketing,* 68(3), 48-62.

Newton, M.A., & Raftery, A.E. (1994). “Approximating Bayesian Inference with the Weighted Likelihood Bootstrap Service and Satisfaction: An Empirical Investigation,” *Journal of the Royal Statistical Society (B)* 56(1) 3-48.

Petrin, A., & Train, K. (2010). “A Control Function Approach to Endogeneity in Consumer Choice Models,” *Journal of Marketing Research,* 47(1) 3-13.

Plummer, M., Best, N., Cowles, K., Vines, K. (2019). Package ‘coda’.

Reynolds, T.J., & Jolly, J.P. (1980). “Measuring Personal Values: An Evaluation of Alternative Methods,” *Journal of Marketing Research,* 17(4) 531-536.

Rossi, P.E., McCulloch, R.E., Allenby, G.M. (1996). “The Value of Purchase History Data in Target Marketing,” *Marketing Science,* 15(4) 321-340.

Roy, R., Chintagunta, P.K., Haldar, S. (1996). “A Framework for Investigating Habits, The Hand of the Past and Heterogeneity in Dynamic Brand Choice,” *Marketing Science*, 15(3) 280-299.

Scott-Morton, F., Zettelmeyer, F., Silva-Risso, J.M. (2001). “Internet Car Retailing,” *The Journal of Industrial Economics,* 49 (4), 501-519.

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

Seetharaman, P B. (2004). “Modeling Multiple Sources of State Dependence in Random Utility Models: A Distributed Lag Approach,” *Marketing Science,* 23(4), 263-271.

Shalizi, R.C., & Thomas, A. (2011). “Homophily and Contagion are Generically Confounded in Observational Social Network Studies,” *Sociological Methods & Research,* 40(2) 211-239.

Sheth, J.N., Newman, B.I., Gross, B.L. (1991). “Why We Buy What We Buy: A Theory of Consumption Values,” *Journal of Business Research,* 22(2) 159-170.

Silva-Risso, J.M. & Ionova, I. (2008). “A Nested Logit Model of Product and Transaction-Type Choice for Planning Automakers’ Pricing and Promotions,” *Marketing Science,* 27(4), 545-566.

Sirgy, M.J. (1982). “Self-Concept in Consumer Behavior: A Critical Review,” *Journal of Consumer Research*, (3) 287-300.

Steel, M. F. (2020). “Model averaging and its use in economics,” *Journal of Economic Literature.*

Theil, H., & Goldberger, A.S. (1961). “On pure and mixed statistical estimation in economics,” *International Economic Review,* 2 65-78.

ter Hofstede, F., Wedel, M., Steenkemp, J.E.M. (2002), “Identifying Spatial Segments in International Markets,” *Marketing Science*, 21(2), 160-177.

van Dijk, A., van Heerde, H.J., Leeflang, P.S.H., Wittink, D.R. (2004). “Similarity-Based Spatial Methods to Estimate Shelf Space Elasticities,” *Quantitative Marketing and Economics*, 2(3), 257-277.

Villas-Boas, J.M., & Winer, R.S. (1999). "Endogeneity in Brand Choice Models," *Management Science,* 45(10) 1324-1338.

Wedel, M. & DeSarbo, W.S. (1994). "A review of recent developments in latent class regression models," *Advanced Methods of Marketing Research, R. Bagozzi (Ed.), Blackwell Pub*, pp.352-388.

Yang, S., & Allenby, G.M. (2003). "Modeling Interdependent Consumer Preferences," *Journal of Marketing Research,* 40(3), 282-294.

Zettelmeyer, F., Scott-Morton, F., Silva-Risso, J.M. (2006). "How the Internet Lowers Prices: Evidence from Matched Survey and Automobile Transaction Data," *Journal of Marketing Research,* 43(2), 168-181.

Table 1. Model Overview

Description	Correlated preferences across consumers	Source of correlation	Method of mixing multiple contiguities	Heterogeneity in consumer response	Previous literature
Model 1: Standard Probit	×	-	NA	×	Common textbook model
Model 2: Random Coefficient Probit	×	-	NA	ZIP code level heterogeneity	Silva-Risso and Ionova (2008)
Model 3: Spatially Correlated Preferences	√	Geographic proximity	NA	×	Duan and Mela (2009), Bezawada et al. (2009) for continuous outcomes. LeSage (2003). Jank and Kannan (2005) for discrete choice.
Model 4: Spatially Correlated Preferences	√	Previous vehicle	NA	×	
Model 5: Spatially Correlated Preferences	√	Geographic proximity + previous vehicle	Direct Mixing of W matrices	×	Yang and Allenby (2003) for binary choice.
Model 6: Structural Uncertainty incorporated Spatially Correlated Preferences	√	Geographic proximity + Previous vehicle	Mixing probabilities from individual W matrices	×	
Model 7: Model 6 + Response Heterogeneity	√	Geographic proximity + Previous vehicle	Mixing probabilities from individual W matrices	ZIP code level heterogeneity	

Table 2. Summary Statistics

Vehicle model	Average price ($)	Average rebate ($)	Average APR (%)	Market share (%)
Accord	21,386	2	6.02	47.66
Altima	21,058	1500	9.16	17.58
Camry	21,208	2	9.57	30.08
Passat	25,332	171	5.16	4.24

Table 3. Model Fit Statistics

Description	Spatial contiguity Matrix	Method of mixing multiple contiguities	Heterogeneity in consumer response	Calibration sample		Holdout sample
				MAD	LMD	MAD
Model 1: Standard Probit	-	NA	-	0.416	– 640.35	0.465
Model 2: Random Coefficient Probit	-	NA	ZIP code level heterogeneity	0.369	– 581.28	0.430
Model 3: Spatially Correlated Preferences	Geographic proximity	NA	-	0.327	– 492.61	0.430
Model 4: Spatially Correlated Preferences	Previous vehicle	NA	-	0.317	– 482.38	0.423
Model 5: Spatially Correlated Preferences	Geographic proximity + previous vehicle	Direct Mixing of W matrices	-	0.325	– 490.52	0.425
Model 6: Structural Uncertainty incorporated Spatially Correlated Preferences	Geographic proximity + Previous vehicle	Mixing probabilities from individual W matrices	-	0.315	– 462.43	0.421
Model 7: Model 6 + Response Heterogeneity	Geographic proximity + Previous vehicle	Mixing probabilities from individual W matrices	ZIP code level heterogeneity	0.304	– 453.84	0.373

Table 4. Parameter Estimates of the Proposed Model (Model 7)

	Posterior Mean	Posterior Std. Dev.
Preference Coefficients		
Accord	**– 4.4749**	1.4357
Altima	**– 2.8943**	1.3233
Camry	– 1.2968	1.3037
Response Coefficients		
Last Make	**2.1791**	1.1802
Price	– 1.5768	0.8818
Price residuals	– 1.1822	0.8253
Closest Dealer	1.1839	1.1887
Spatial Coefficients: Marginal Distribution		
Accord	0.0004	0.0517
Altima	– 0.0013	0.0353
Camry	– 0.0025	0.0372
Spatial Correlation Parameters		
Spillover of Geographic preferences (ρ_G)	**0.5804**	0.1192
Spillover of Vehicular preferences (ρ_V)	**0.3239**	0.0864
Mixing Parameter		
Representation of Geographic Information	**0.0454**	0.0166

Figure 1. Schematic Representation of Proposed Model

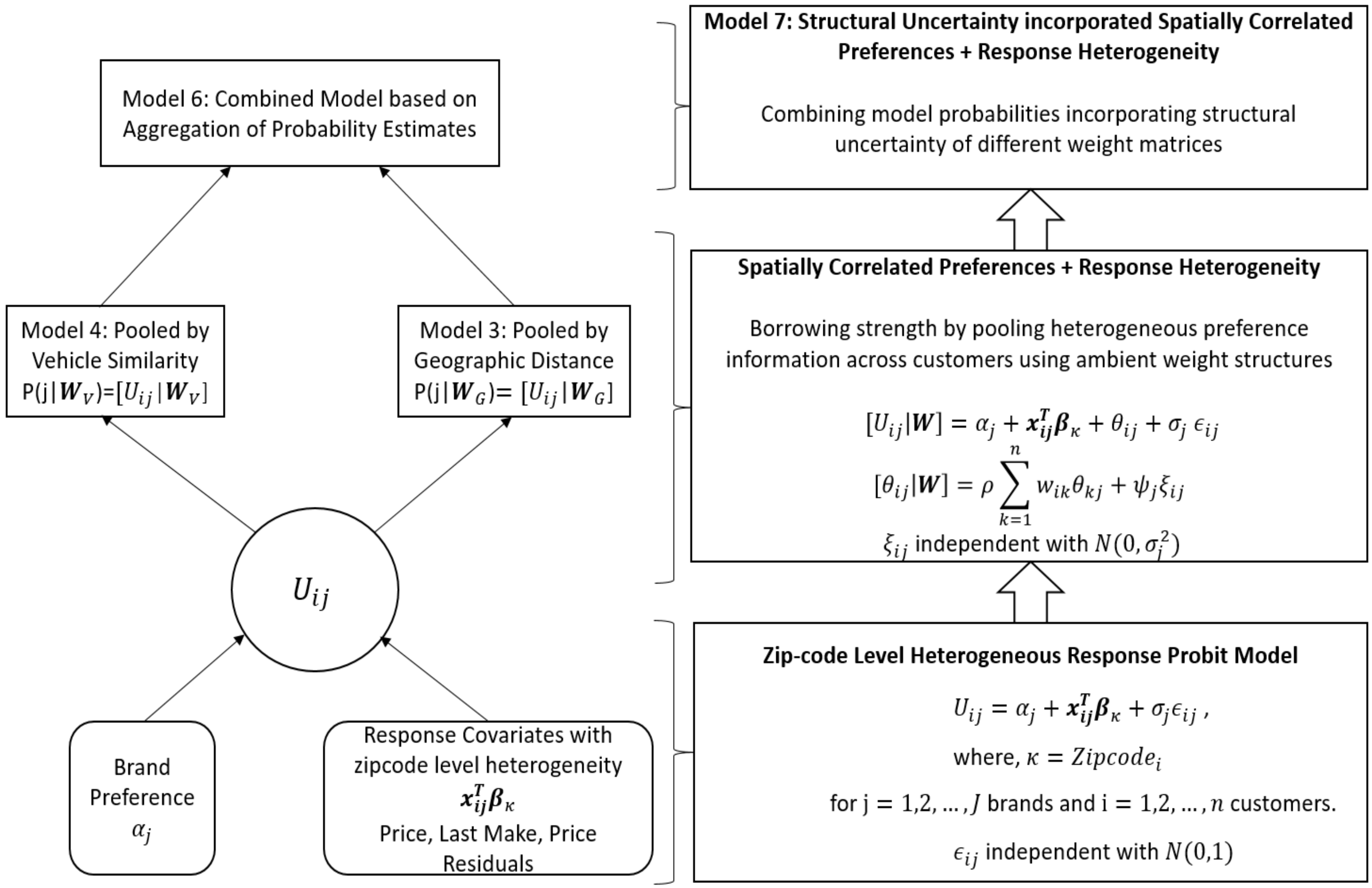
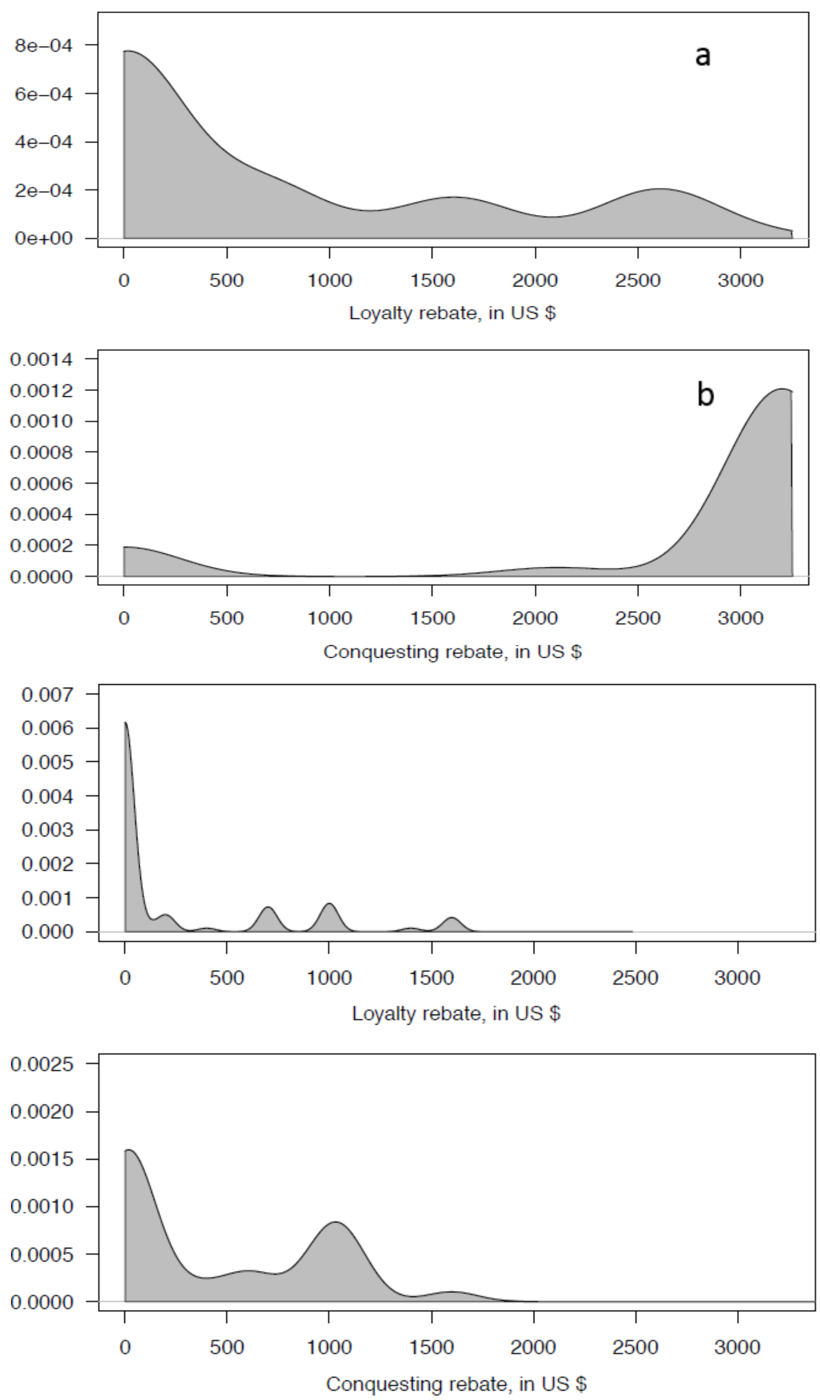


Figure 2. Histograms of Optimal ZIP code-Specific Loyalty and Conquesting Rebates for Model 2 (panels a and b) and Model 7 (panels c and d). The horizontal axis in each plot shows the optimal rebate amount and the vertical axis shows the density value.



Appendix: MCMC Algorithms

MCMC update for Model 3, 4 and 5.

0. Set $WP = \lambda^{(0)} W_G + \big(1 - \lambda^{(0)}\big) W_V$.

1. $\underline{u}_i \,|\, y_i, \beta^{(0)}, \theta^{(0)}, \Sigma^{(0)}, \Psi^{(0)}, \rho^{(0)}, \lambda^{(0)} \sim$ Truncated Normal$\big(X_i \beta_i^{(0)} + \underline{\theta}_i^{(0)}, \Sigma^{(0)}\big)$, with the truncation that $u_{ij} > u_{ij'}$ if $y_i = j$. Here $\underline{u}_i = \big(u_{(i-1)J+1}, \ldots, u_{(i-1)J+J}\big)$, and $\underline{\theta}_i^{(0)} = \big(\theta_{(i-1)J+1}^{(0)}, \ldots, \theta_{(i-1)J+J}^{(0)}\big)$.

2. $\beta^{(1)} \,|\, u, y_i, \beta^{(0)}, \theta^{(0)}, \Sigma^{(0)}, \Psi^{(0)}, \rho^{(0)}, \lambda^{(0)} \sim \text{Normal}\big(\widetilde{\beta}, \Phi\big)$ where

$$\Phi = \left[X^\top ({\Sigma^{(0)}}^{-1} \otimes I_n) X\right]^{-1}, \text{ and}$$

$$\widetilde{\beta} = \Phi \left[X^\top ({\Sigma^{(0)}}^{-1} \otimes I_n)(u - \theta^{(0)})\right].$$

3. $\theta^{(1)} \,|\, u, y_i, \beta^{(1)}, \theta^{(0)}, \Sigma^{(0)}, \Psi^{(0)}, \rho^{(0)}, \lambda^{(0)} \sim \text{Normal}(Qe, Q)$, where $e = u - X\beta^{(1)}$,

$$R = \big(I_n - \rho^{(0)} WP \otimes I_J\big), \quad \text{and} \quad Q = \left({\Sigma^{(0)}}^{-1} \otimes I_n - {\Phi^{(0)}}^{-1} R^\top R\right).$$

4. ${\sigma_j^{(1)}}^2 \,|\, u, y_i, \beta^{(1)}, \theta^{(1)}, \Sigma^{(0)}, \Psi^{(0)}, \rho^{(0)}, \lambda^{(0)} \sim \text{InvGamma}\big(a_\sigma, b_\sigma + {\underline{e}'_j}^\top \underline{e}'_j\big)$, where $a_\sigma = n/2 + 2$, $b_\sigma = 10$,

$$e' = u - \theta^{(1)} - X\beta^{(1)}, \quad \text{and} \quad \underline{e}'_j = \big(e'_j, e'_{J+j}, \ldots, e'_{(n-1)J+j}\big)^\top.$$

5. ${\psi_j^{(1)}}^2 \,|\, u, y_i, \beta^{(1)}, \theta^{(1)}, \Sigma^{(1)}, \Psi^{(0)}, \rho^{(0)}, \lambda^{(0)} \sim \text{InvGamma}\big(a_\psi, b_\psi + \zeta_j^\top \zeta_j\big)$, where $a_\psi = n/2 + 2$, $b_\psi = 10$,

$$\underline{\theta}_j = \left(\theta_j^{(1)}, \theta_{J+j}^{(1)}, \ldots, \theta_{(n-1)J+j}^{(1)}\right)^\top, \quad \text{and} \quad \zeta_j = \big(I_n - \rho^{(0)} WP\big)\underline{\theta}_j.$$

6. $\rho^{(1)} \,|\, u, y_i, \beta^{(1)}, \theta^{(1)}, \Sigma^{(1)}, \Psi^{(1)}, \rho^{(0)}, \lambda^{(0)}$ is equal to $\widetilde{\rho}$ with probability π and equal to $\rho^{(0)}$ with probability $1 - \pi$, where

$$\widetilde{\rho} \sim \text{Normal}\big(\rho^{(0)}, 0.005^2\big),$$

$$\pi = \min\left\{1, \frac{|\widetilde{R}| \exp\Big(-0.5\psi_j^{(1)^{-2}}\underline{\theta}_j^\top \widetilde{R}^\top \widetilde{R}\underline{\theta}_j\Big)}{|R^{(0)}| \exp\Big(-0.5\psi_j^{(1)^{-2}}\underline{\theta}_j^\top R^{(0)\top} R^{(0)}\underline{\theta}_j\Big)}\right\} \text{ with}$$

$$R^{(0)} = I_n - \rho^{(0)}WP, \quad \text{and} \quad \widetilde{R} = I_n - \widetilde{\rho}WP.$$

7. $\lambda^{(1)} \mid u, y_i, \beta^{(1)}, \theta^{(1)}, \Sigma^{(1)}, \Psi^{(1)}, \rho^{(1)}, \lambda^{(0)}$ is equal to $\widetilde{\lambda}$ with probability π and equal to $\lambda^{(0)}$ with probability $1 - \pi$, where

$$\widetilde{\lambda} = 1/\{1 - \exp(-\widetilde{\alpha})\} \text{ with } \widetilde{\alpha} \sim \text{Normal}\Big(\log(\lambda^{(0)}/(1 + \log\lambda^{(0)}), 0.005^2\Big), \text{ and}$$

$$\pi = \min\left\{1, \frac{|\widetilde{R}| \exp\Big(-0.5\psi_j^{(1)^{-2}}\underline{\theta}_j^\top \widetilde{R}^\top \widetilde{R}\underline{\theta}_j\Big)}{|R^{(1)}| \exp\Big(-0.5\psi_j^{(1)^{-2}}\underline{\theta}_j^\top R^{(1)\top} R^{(1)}\underline{\theta}_j\Big)}\right\} \text{ with}$$

$$R^{(1)} = \big(I_n - \rho^{(1)}WP\big), \widetilde{R} = \big(I_n - \rho^{(1)}\widetilde{WP}\big), \text{ and } \widetilde{WP} = \widetilde{\lambda}W_G + \big(1 - \widetilde{\lambda}\big)W_V.$$

*For model 3 and 4, skip step 7, and set $\lambda^{(1)} = \lambda^{(0)} = 1$ and $\lambda^{(1)} = \lambda^{(0)} = 0$ respectively.

Remark on model estimation. The model parameters σ_j and ψ_j are not separately identifiable. But our purpose of the proposed model is in predicting the discrete choices of the customers. Thus we can assess the predictive performance of the model by restricting the ratios of σ_j/ψ_j in the compact set $[0.10, 10]$. Our MCMC strategy for model estimation implemented this restriction. Additionally, the response coefficients βs are identifiable in the proposed model and thus we can provide statistical inference for these parameters of interest based on the MCMC draws of these parameters (as discussed in Section 4.1 of the paper).

MCMC Update for Model 6 and 7.

Let C_i is a categorical variable with possible values in $\{c_1, \ldots, c_L\}$. In the paper they denote the whether in the structural uncertainty model the individual follows geographic similarity, or follows vehicle similarity based utility model. Then $[U_{ij} \mid C_i = c_l] = [U_{ij} \mid W^{(l)}]$, where $W^{(1)}, \ldots, W^{(L)}$ are the L possible spatial similarity matrices, and $[U_{ij} \mid W^{(l)}]$ specifies a preference model. To generalize our notation we add (l) in the superscript, e.g., σ_j becomes $\sigma_j^{(l)}$. Let $\gamma_l = \Pr(C_i = c_l)$

for $l = 1, \ldots, L$.

Priors: The corresponding parameters from separate components have independent and identical prior distributions. Thus, we only need to specify the additional prior for the γ_l's. We let $(\gamma_1, \ldots, \gamma_L)$ ~Dirichlet$(3/L, \ldots, 3/L)$. In particular, when $L = 2$, $\gamma := \gamma_1 \sim$ Beta$(3, 3)$.

MCMC for $(\theta_{ij}^{(l)}, \beta_k^{(l)}, \rho^{(l)}, \psi_j^{(l)}, \sigma_j^{(l)}, \gamma_l)$*:* The index i runs from 1 through n, j runs from 1 through $J-1$ (where J is the number of choices in a multiple choice model), k runs from 1 through K the number of covariates, and l runs from 1 through L. We summarize here the MCMC step for updating the model parameters from $\zeta^{0(l)} := \left(\theta_{ij}^{0(l)}, \beta_k^{0(l)}, \rho^{0(l)}, \psi_j^{0(l)}, \sigma_j^{0(l)}, \gamma_l^0, C_i^0\right)$ to $\zeta^{1(l)} = \left(\theta_{ij}^{1(l)}, \beta_k^{1(l)}, \rho^{1(l)}, \psi_j^{1(l)}, \sigma_j^{1(l)}, \gamma_l^1, C_i^1\right)$.

1. *Update component labels:*

   i. For each j, simulate $\left(\widetilde{\theta}_{1j}^{(l)}, \ldots, \widetilde{\theta}_{nj}^{(l)}\right)$ from Normal$(\mathbf{0}, \Xi^{(l)})$, where $\Xi^{(l)} = \left\{I - 0.4(W^{(l)T} + W^{(l)}) - 0.4^2 W^{(l)T} W^{(l)}\right\}^{-1}$.

   ii. Calculate $\tilde{p}_i^{(l)} = \Pr\left(y_i \mid W^{(l)}, \widetilde{\theta}_{ij}^{0(l)}, \beta_k^{0(l)}, \rho^{0(l)}, \psi_j^{0(l)}, \sigma_j^{0(l)}\right)$.

   iii. Calculate $\hat{p}_i^{(l)} = \Pr\left(y_i \mid W^{(l)}, \zeta^{0(l)}\right)$ for $C_i^0 = c_l$. Set $p_i^{(l)} = \hat{p}_i^{(l)} 1\left(C_i^0 = c_l\right) + \tilde{p}_i^{(l)} 1\left(C_i^0 \neq c_l\right)$.

   iv. Update $C_i^1 = c_l$ with probability proportional to $\gamma_l^0 p_i^{(l)}$.

2. *Update the component specific parameters:*

   For each l, update $(\theta_{ij}^{1(l)}, \beta_k^{1(l)}, \rho^{1(l)}, \psi_j^{1(l)}, \sigma_j^{1(l)})$, using the data where $C_i^1 = c_l$. The details are same as Steps 1 through 6 detailed for Model 5, with two changes: (a) the spatial similarity matrix is $W^{(l)}$ in place of WP there, and (b) only the subset of data where $C_i^1 = c_l$ are used. Also, for Model 7, update the response coefficient β as in a random coefficients model for the three digit zip codes.

3. *Update* γ_l*:* $\left(\gamma_1^1, \ldots, \gamma_L^1\right) \sim$ Dirichlet$\left(3 + \sum_{i=1}^n 1\left(C_i^1 = c_1\right), \ldots, 3 + \sum_{i=1}^n 1\left(C_i^1 = c_L\right)\right)$.

Computational details of implementation of the MCMC algorithms and convergence check. Here we give some further remarks about the implementation of these algorithms, their computation, and verification. These algorithms were implemented on the `R` software. Some existing packages such as `biglm` and `bayesm` were used and `Rcpp` package was used for implementing various matrix computations. Further, parts of the code were parallelized using the `foreach` and `doSNOW` package in 4 cpu cores. For each model we verified convergences using various statistics trace plot that are standard in the packages `coda` and `mcmcplot`.

Alternative $J = 4$ was the base model whose utility is reset to 0, and variance terms were scaled by the normalizing $\sigma_1 = 1$. These implementations were run on a 1.7GHz quad-core Intel i7 processor with 32GB ram. Model 3–5 took about 30 hours to draw 100000 mcmc samples and Model 6 and 7 took about 50 hours to draw 100000 mcmc samples.