Journal of Mathematical Psychology 82 (2018) 56–72

Contents lists available at ScienceDirect

Journal of Mathematical Psychology

journal homepage: www.elsevier.com/locate/jmp

A hierarchical Bayesian approach for examining heterogeneity in choice decisions

Sunghoon Kim[a,*], Wayne S. DeSarbo[b], Duncan K.H. Fong[b]

[a] Department of Marketing, W.P. Carey School of Business, Arizona State University, Tempe, AZ, 85258, United States
[b] Department of Marketing, Smeal College of Business, Pennsylvania State University, University Park, PA, 16802, United States

HIGHLIGHTS

- A new hierarchical Bayes multivariate probit mixture model incorporating variable selection.
- Heterogeneity of the important features that drive consumer choices.
- A consumer psychology application involving the consideration of Sports Utility vehicles.
- Favorable methodological comparisons with a variety of alternative benchmark models.

ARTICLE INFO

Article history:
Received 9 April 2017
Received in revised form 22 October 2017

Keywords:
Finite mixtures
Variable Selection
Bayesian multivariate probit models
Consumer psychology
Choice heterogeneity

ABSTRACT

There is a vast behavioral decision theory literature that suggests different individuals may utilize and/or weigh different attributes of an object to form the basis of their opinions, attitudes, choices, and/or evaluations of such stimuli. This heterogeneity of information utilization and importance can be due to several different factors such as differing goals, level of expertise, contextual factors, knowledge accessibility, time pressure, involvement, mood states, task complexity, communication or influence of relevant others, etc. This phenomenon is particularly pertinent to the evaluation of stimuli involving large numbers of underlying attributes or features. We propose a new hierarchical Bayesian multivariate probit mixture model with variable selection accommodating such forms of choice heterogeneity. Based on a Monte Carlo simulation study, we demonstrate that the proposed model can successfully recover true parameters in a robust manner. Next, we provide a consumer psychology application involving consideration to buy choices for intended consumers of large Sports Utility Vehicles. The application illustrates that the proposed model outperforms several comparison benchmark choice models with respect to face validity and choice predictive validation performance.

1. Introduction

There is a plethora of behavioral decision theory research in psychology that suggests that individuals may utilize and/or weigh different attributes of stimuli to form their opinions, attitudes, and/or evaluations (e.g., preference or choice). As recently suggested by Park, Rajagopel, Dillon, and Chaiy (2017), this heterogeneity of information utilization, importance, and process can be due to several different factors including individuals' differing goals (Payne, Bettman, & Johnson, 1993), individual differences involving prior knowledge and/or level of expertise (Alba & Hutchinson, 1987), contextual factors (Hutchinson & Alba, 1991), knowledge accessibility (Feldman & Lynch, 1988), time pressure (Wright & Weitz, 1977), level of involvement (Petty & Cacioppo, 2012), mood states (Isen, 1993), task complexity (DeSarbo, Fong, Liechty, & Coupland, 2005), etc. In consumer psychology, many researchers such as Payne, Bettman, and Luce (1996) have summarized several well documented non-compensatory decision heuristics (e.g., elimination by aspects, lexicographic rules, satisficing strategies, etc.) that permit consumers to simplify their complicated decision tasks by selectively processing limited attribute/feature information especially in consumptive situations where large numbers of attributes or features are involved. Related to the choice consideration-set heuristics, advances in recent methods (e.g., Bayesian inferences, machine-learning, etc.) help to understand a variety of heuristics including applications in modern complex products with large number of features (Hauser, 2014). Kamakura, Kim, and Lee (1996) identify both preference and structural heterogeneity in choice which purports potential heterogeneity in both the parameters of the choice model being

* Corresponding author.
E-mail addresses: skim348@asu.edu (S. Kim), wsd6@psu.edu (W.S. DeSarbo), i2v@psu.edu (D.K.H. Fong).

https://doi.org/10.1016/j.jmp.2017.11.002

estimated, as well as potential heterogeneity in the choice process undertaken. Okada and Lee (2016) and Park, DeSarbo, and Liechty (2008) have examined both forms of such heterogeneity in various types of multidimensional scaling models.

Behavioral research concerning individuals' goals has found that decision makers with accuracy goals undertake more extensive processing of attribute information than decision makers with effort minimization goals (Payne et al., 1993); thus, individuals with the goal of effort minimization are likely to utilize fewer attributes than individuals with the goal of accuracy in decision-making. Similarly, individuals with justification goals have been shown to focus on attributes that help them justify their final choices (Kunda, 1990). Hence, the type of goal can be a very important determinant of the number and type of attributes utilized during decision making and choice.

In a similar vein, the existent research on expertise has shown that experts possess greater and more detailed knowledge structures about categories than novices (Johnson & Mervis, 1997), are able to recall more dimensions/attributes about alternatives than novices (Vicente & Wang, 1998), and are able to make more accurate attribute-benefits linkages than novices (Dellaert & Stremersch, 2005). This suggests that when decision making is memory-based, experts will utilize more attributes than novices, and are more likely to focus on important and relevant attributes; likewise, novices are likely to rely on more salient and prototypical attributes/dimensions (Alba & Hutchinson, 1987). Cowley and Mitchell (2003) found that novices who were exposed to information about a consumer product in the context of a specific use situation were less able to recall the product in the context of a different use situation; thus, their usage of product information was related to use situation. However, experts were better able to reorganize product information in memory and recall product information that was appropriate for new use situations (Hutchinson & Alba, 1991; Miller & Ginter, 1979). This also suggests the possibility of different choice processes being employed to make choice decisions.

Apart from goals and expertise/knowledge, contextual variables such as time pressure, mood state, and involvement can also affect attribute/feature information selection and utilization. For example, under moderate time pressure, individuals are likely to process each stimulus alternative separately; while under severe time pressure, individuals have been shown to switch to select a few important attributes and evaluate alternatives based on this restricted set of dimensions (Payne et al., 1996). Houston and Sherman (1995) found that the starting alternative in a choice process determined the type of attribute that received greater weight during choice. Thus, features shared by the choice alternatives were canceled and greater weight was placed on the unique features of the alternative that was the starting point for comparison. Since the starting alternative in a choice set is likely to be different for different individuals, different attributes would emerge as unique versus common resulting in different attributes being weighed differently during the choice process. Involvement with the stimulus category or the decision has also been shown to have a significant effect on various decision-making processes and information processing. Individuals with higher levels of involvement with the decision or the object have been shown to pay greater attention to the decision-making task, and process more information than individuals with lower levels of involvement. Highly involved decision makers have also been shown to focus on relevant aspects of the choice task as compared to subjects with low involvement levels who tend to focus on peripheral aspects of the choice task (Petty & Cacioppo, 2012; Petty, Cacioppo, & Goldman, 1981; Sujan, 1985). Finally, research on positive affect has shown that people in a positive mood are cognitively more flexible than people in negative or neutral moods, and have been shown to be able to utilize more attributes and broader dimensions during decision making (e.g., Isen, 1993; Isen, Daubman, & Nowicki, 1987).

Thus, there can be substantial heterogeneity in the way different subjects derive their preferences, choices, and/or decisions in terms of the different types of attribute information focused upon, as well as the choice rules employed. This heterogeneity becomes particularly relevant when dealing with their evaluations of stimuli defined on many attributes or features and with subjects with differing level of familiarity or expertise with the study stimuli. Let's now explore how this heterogeneity in information utilization manifests itself in a typical choice setting. Consider a binary choice case where each subject must choose between selecting or not-selecting a certain stimulus (in our consumer psychology application, we use intended consumers' decisions to consider buying or not buying brands in a designated product class) described by different combinations of P known attributes or features; and, also assume each subject repeat this choice for M different alternatives. Let $\underline{C}_i = (C_{i1}, \ldots, C_{iM})^T$ denote the M choices made by the ith subject which depends upon stimulus features via the following individual-level generalized linear model:

$$\underline{C}_i = g\left(\underline{X}_i \underline{\beta}_i + \underline{\epsilon}_i\right), i = 1, \ldots, n,$$

where $\underline{\epsilon}_i$ denotes the error term and $\underline{X}_i$ denotes the corresponding $M \times P$ attribute matrix. Because the M choices made by each subject are likely to be interrelated, we propose a multivariate probit model framework for this setting. Regarding the multivariate setting, recent empirical findings in psychology demonstrate that previous choice tasks can affect the current choice, indicating inter-dependence across multiple choices (Leong & Hensher, 2012). We employ a finite mixture formulation to parsimoniously reflect subject response (choice) heterogeneity (see Rossi, Allenby, & McCulloch, 2005 for a review of alternative approaches utilizing hierarchical Bayesian formulations of this heterogeneous choice model; see also Wedel et al., 1999 for developments in Marketing, Bhat, 2017 in engineering, Li & Ansari, 2014 in economics, Yang, 2005 in transportation, etc.). As to be developed shortly, we also provide for simultaneous variable selection per derived cluster.

Although finite mixture based (multivariate) choice models have been previously formulated (see Arminger, Clogg, & Cheng, 2000; Bontemps & Toussile, 2013; Wedel & Kamakura, 2000) and applied to several types of applications (e.g., in revealed choice conjoint analysis), there are several potentially problematic issues to resolve. One major challenge concerns the increasing number of stimulus attributes/features encountered in many applications (i.e., large P). For example, many manufacturers are currently adding more and more features into consumer products such as smart phones, digital cameras, flat screen TV's, automobiles, laptop computers, tablets, etc. This trend restricts the use of some traditional methods such as conjoint analysis or revealed preference analysis where it is recommended not to utilize more than six or seven attributes or features (Green & Srinivasan, 1978) to collect such preferences or choices. In responding to this new challenge, researchers in choice modeling have proposed some recent alternatives including hybrid techniques using self-explication (Johnson, 1987) methods that rely on subjects in the experiment to reveal explicitly what the important features are to them. These alternative approaches, however, do not often reflect real-life choice scenarios and may therefore provide results that are incomplete and/or difficult to interpret. Given a long list of attributes/features, subjects usually make their choice decisions only based on a subset of important attributes/features due to convenience, cost of thinking, or lack of expertise about some attributes/features, etc. (Gilbride, Allenby, & Brazell, 2006), in contrast to the assumption held in traditional choice methods that subjects consider every attribute/feature before manifesting their choice. Given the vast behavioral research cited earlier, such heterogeneity in attribute/feature selection should be incorporated into such choice models as well. Another challenge comes from the restriction that the number of choices, M, cannot be very large before the subject gets bored or fatigued. When the dimension P gets large, each individual level model above can suffer from the "high dimension, low sample-size" paradigm which makes inference on attribute/feature selection for each individual model even more challenging.

2. Literature review

Regarding the general clustering model with variable selection, Ritter (2014) has recently documented the need for variable selection in the use of cluster analysis in his review of the classification literature over the past several decades. Recent clustering models have begun to utilize stochastic variable selection techniques: the Bayesian variable selection approach (George & McCulloch, 1993, 1997) and Lasso (Tibshirani, 1996) or SCAD (Fan & Li, 2001). Tadesse, Sha, and Vannucci (2005) propose a Bayesian variable selection model in clustering high-dimensional data focusing on data which have smaller sample size than the number of covariates. Kim, Tadesse, and Vannucci (2006) extend the Tadesse et al.'s model (2005) by formulating the clustering via Dirichlet process mixture models. In addition, Raftery and Dean (2006) have proposed an alternative variable selection approach in clustering which provides a search algorithm and adopts an approximate Bayes factor for finding a local optimum (see also Dean and Raftery, 2010). Recently, White, Wyse, and Murphy (2016) suggest a model-based clustering approach for multivariate categorical data using Bayesian variable selection. However, these Bayesian models have been developed in the setting without dependent variables, and do not solve the problem of cluster-wise heterogeneity in independent variable selection.

Ghosh, Herring, and Siega-Riz (2011) had proposed a latent class logistic regression model using a Bayesian variable selection technique. In a related penalized likelihood approach, Khalili and Chen (2007) proposed a finite mixture regression model for different response dependent variable response scales by employing the Lasso or SCAD technique for variable selection. More recently, Andrews and McNicholas (2014) proposed another variable selection technique based on within-group variance under a model-based clustering framework. Bayesian mixture of lasso regressions with particle MCMC was proposed by Cozzini, Jasra, Montana, and Persing (2014), and Benati and García (2014) proposed a model for discarding masking variables where the criterion was the minimization of the total distance l_1. In addition, Bouveyron and Brunet-Saumard (2014) suggested a clustering model for selecting discriminative variables by using sparsity in the loading matrix of the Fisher-EM algorithm. However, these approaches do not consider the multivariate setting of multiple or repeated choices (in a clustering context) which is commonly used in a variety of applications (see DeSoete and DeSarbo, 1991; Kim, Allenby, and Rossi, 2002; Dubé, 2004; Li, Sun, and Wilcox, 2005).

To understand the relationship between M choices, we propose a Bayesian multivariate probit model framework (see Liu and Daniels, 2006) where the model estimates cluster-wise regression coefficients and correlation matrices. To deal with the estimation of a correlation matrix, Chib and Greenberg (1998) assume that the off-diagonal components of the correlation matrix follow a multivariate normal distribution truncated to the space of correlation matrix and use the Metropolis–Hastings algorithm to sample random deviates from the corresponding posterior distribution. But, the sampled values are not guaranteed to be positive definite, and the model can have a problem of slow mixing. The method suggested by Liechty, Liechty, and Müller (2004) also employs the Metropolis–Hastings algorithm, but their method is computationally intensive since it simulates all the components of the correlation matrix one by one through a latent variable. In the proposed method, we employ a parameter expansion technique (Liu & Daniels, 2006; Liu & Wu, 1999) to resolve the computational challenges for generating random correlation matrices as we describe in more detail shortly.

To address all pre-described issues, we propose a new Bayesian multivariate probit mixture model simultaneously accommodating stochastic variable selection for searching key independent variables by derived cluster as well as estimating cluster-wise correlation matrices by using a parameter expansion technique. We note that none of the previous models can simultaneously conduct multivariate choice (binary) regression analysis for the correlated choices as well as identify cluster level key drivers. Thus, the proposed model can be a powerful tool for researchers to examine heterogeneity in choice decisions.

3. The proposed heterogeneous Bayesian multivariate probit mixture model with variable selection

3.1. The model and priors

For our proposed multivariate probit mixture model, we assume that each subject has provided M multiple binary responses, and we let $\underline{c}_i = (c_{i1}, \ldots, c_{iM})^T$ denote the M dimensional vector of observed binary 0 or 1 choice responses for the ith subject, $i = 1, \ldots, n$. Let $\underline{Y}_i = (Y_{i1}, \ldots, Y_{iM})^T$ denote a vector of latent utilities such that:

$$c_{im} = \begin{cases} 1, & if\ Y_{im} > 0 \\ 0, & otherwise, \end{cases} \quad m = 1, \ldots, M. \tag{1}$$

Suppose $\underline{X}_i$ is the corresponding $M \times P$ design or attribute matrix (including an intercept term); we assume an underlying multivariate regression model for the latent utilities that allows us to simultaneously perform clustering:

$$\underline{Y}_i = \underline{X}_i \underline{\beta}_{G_i} + \underline{\varepsilon}_i, i = 1, \ldots., n, \tag{2}$$

where $\underline{\beta}_{G_i}$ is a $P \times 1$ vector of cluster level coefficients, G_i is a latent variable indicating cluster membership, where $G_i = k$ means that subject i is in cluster $k \in \{1, \ldots, K\}$. In addition, $\underline{\varepsilon}_i$ is a $M \times 1$ vector of residual errors normally distributed as $N\left(\underline{\mathbf{0}}, \underline{R}_{G_i}\right)$, where $\underline{R}_k$ is the $M \times M$ error covariance matrix in cluster k. For identification purposes, the diagonal elements of the error covariance matrix $\underline{R}_k$, $k \in \{1, \ldots, K\}$, are assigned the value of 1 so that $\underline{R}_k$ becomes a correlation matrix (see Chib & Greenberg (1998) for the case of the multivariate probit model). Then, we assume vague priors for the correlation matrices which are independent and locally uniform priors:

$$\pi_1\left(\underline{R}_k\right) \propto 1, \ \underline{R}_k \in \mathcal{J}, \quad k = 1, \ldots, K, \tag{3}$$

where the parameter space $\mathcal{J}$ is convex and compact. Note that such priors are vague but proper; see Rousseeuw and Molenberghs (1994).

For the cluster indicator variable, G_i is assumed to follow a categorical distribution with positive parameters, $d_1, \ldots, d_K$ (i.e., they are probabilities of $G_i = k$, $k = 1, \ldots, K$):

$$\pi\left(G_i | \underline{d}\right) = Categorical\left(d_1, \ldots, d_K\right), \tag{4}$$

and $\underline{d} = (d_1, \ldots, d_K)$ follows a Dirichlet distribution, *Dirichlet* $(\alpha_1, \ldots, \alpha_K)$. To perform variable selection, we assume a spike and slab prior for the components of $\underline{\beta}_k$ (see Ishwaran and Rao, 2011; Kim, Fong, and DeSarbo, 2012; Mitchell and Beauchamp, 1988):

$$\pi\left(\underline{\beta}_k | Z_{kp}, p = 1, \ldots, P\right) = \prod_{p=1}^{P}\left[Z_{kp} N\left(\beta_{kp}|0, \tau_p^2\right) + \left(1 - Z_{kp}\right)\mathbf{1}_{\{\beta_{kp}=0\}}\right], \quad (5)$$

where Z_{kp} is a Bernoulli random variable with a parameter w and $\mathbf{1}_{\{*\}}$ is the indicator function. As typical, for τ_p^2 and w, we assume:

$$\pi(w) = Beta(a, b); \quad (6)$$

$$\pi\left(\tau_p^2\right) = InvGamma\left(s_{p1}, s_{p2}\right). \quad (7)$$

3.2. The proposed Markov Chain Monte Carlo algorithm

In Bayesian computation, it is important to generate an approximate sample from the joint posterior distribution which can be achieved by drawing random deviates iteratively and recursively from the full conditional distributions. However, the Markov Chain Monte Carlo (MCMC) algorithm will break down here when generating random deviates iteratively and recursively from the full conditional distributions of β_{kp} and Z_{kp}. To tackle the computational problem, we employ a partially collapsed Gibbs sampling strategy (Van Dyk & Park, 2008) to generate β_{kp} from its full conditional distribution, but draw Z_{kp} from a distribution of Z_{kp} which is not conditional on β_{kp} (The derivation of all required conditional distributions are given in Appendix A):

$$\pi\left(Y_{im}|\, allothers\right)\ \textit{is a Truncated normal distribution}, \quad i = 1, \ldots, n, m = 1, \ldots, M. \quad (8)$$

$$\pi\left(G_i|\ all\ others\right)\ \textit{is a Categorical distribution}. \quad (9)$$

$$\pi\left(\underline{d}|\ all\ others\right)\ \textit{is a Dirichlet distribution}. \quad (10)$$

$$\pi\left(\beta_{kp} = 0|Z_{kp} = 0\ and\ others\right) = 1 \quad \text{and} \quad (11)$$

$$\pi\left(\beta_{kp}|Z_{kp} = 1\ and\ others\right)\ \textit{is a Normal distribution}, \quad k = 1, \ldots, K \text{ and } p = 1, \ldots, P.$$

$$\pi\left(Z_{kp}|\ all\ others\ except\, \beta_{kp}\right)\ \textit{is a Bernoulli distribution}. \quad (12)$$

$$\pi\left(w|\ all\ others\right)\ \textit{is a Beta distribution}. \quad (13)$$

$$\pi\left(\tau_p^2|\ all\ others\right)\ \textit{is an inverse Gamma distribution}. \quad (14)$$

The full conditional distribution of the correlation matrix $\underline{R}_k$ is not a standard probability distribution. To speed up the MCMC chain for generating the correlation matrices across clusters, we employ a parameter expansion technique (Liu & Daniels, 2006; Liu & Wu, 1999) through the following one-to-one transformation mapping from $\{\underline{Y}_i, \underline{R}_k\}$ to $\{\underline{W}_i, \underline{\Sigma}_k\}$ $(i \in id_k)$:

$$\begin{cases} \underline{Y}_i = \underline{X}_i \underline{\beta}_{G_i} + \underline{D}_{G_i}^{-1}\underline{W}_i \\ \underline{R}_k = \underline{D}_k^{-1}\underline{\Sigma}_k \underline{D}_k^{-1}, \end{cases} \quad (15)$$

where $\underline{\Sigma}_k$ is a $M \times M$ positive definite matrix, $\sum_{i\in id_k} W_{im}^2 = 1$ for $m = 1, \ldots, M$ ($id_k = \{i : G_i = k\}$), and $\underline{D}_k$ is a diagonal matrix. Then, with an appropriate candidate prior for $\underline{R}_k$ (see Appendix A for details), the full conditional distribution of $\underline{\Sigma}_k$ follows an Inverse Wishart distribution:

$$\underline{\Sigma}_k \sim \text{InvWishart}\left(\upsilon_k, \underline{S}_k\right), \quad (16)$$

where $\upsilon_k = n_k$ ($n_k = \sum_{i=1}^{n}\mathbf{1}_{\{G_i=k\}}$), $\underline{S}_k = \left(\sum_{i\in id_k}\underline{W}_i\underline{W}_i^T\right)$, and $\underline{W}_k = \underline{D}_k\left(\underline{Y}_{[id_k]} - \underline{X}_{id_k}\underline{\beta}_{G_i}\right)$. Next, we translate $\underline{\Sigma}_k$ into $\underline{R}_k$, and accept the candidate $\underline{R}_k$ value using a Metropolis–Hastings step with acceptance rate: $\alpha\ \left(= \min\left\{1, \exp\left\{\frac{M+1}{2}\left(\log\left|\underline{R}_k\right| - \log\left|\underline{R}_k^l\right|\right)\right\}\right\}\right)$ at iteration $l + 1$.

3.3. Technical issues

We employ the mixture adapted BIC and DIC for model selection in terms of determining a parsimonious value of K, the number of clusters (Celeux, Forbes, Robert, & Titterington, 2006; Spiegelhalter, Best, Carlin, & Van Der Linde, 2002; Steele & Raftery, 2010). Regarding the decision criteria of variable selection, the ratio of the posterior odds and prior odds is adopted. Here, posterior odds $= \frac{Number\ of\ Z=1\ in\ MCMC\ Samples}{Number\ of\ Z=0\ in\ MCMC\ Samples}$, and prior odds $= \frac{E[P(Z_{kp}=1)]}{E[P(Z_{kp}=0)]} = \frac{E(w)}{E(1-w)} = \frac{a/(a+b)}{b/(a+b)} = \frac{a}{b}$, where a and b are hyperparameters of the prior distribution for w. Hence, the odds ratio is given by: $\frac{Number\ of\ Z=1\ in\ MCMC\ Samples}{Number\ of\ Z=0\ in\ MCMC\ Samples} \times \frac{b}{a}$, and a cutoff value of 20 is used (Jeffreys, 1961), in contrast to the 5% significance level used for traditional hypothesis testing.

Finally, to address the well-known label switching problem when using MCMC for Bayesian mixture models, we follow Marin and Robert (2007) and Kim et al. (2012) to first simulate from the unconstrained posterior distribution, and then impose identifiability constraints on the generated MCMC sample. Specifically, once the simulation is completed, the $\underline{\beta}_k$ are relabeled for each MCMC iteration according to the constraint: $\beta_{1p} < \beta_{2p} < \cdots < \beta_{Kp}$ for a given component p. Then, other parameters $\underline{Z}$, $\underline{G}$, $\underline{d}$ and $\underline{R}$ are reordered accordingly to match that of $\underline{\beta}_k$ (see also Sperrin, Jaki, and Wit, 2010). Because the proposed Bayesian finite mixture model contains many parameters, this simple algorithm is especially beneficial in terms of computational time savings.

4. A Monte Carlo simulation study

To test the performance of the proposed Bayesian multivariate probit mixture model with variable selection, we conducted a small Monte Carlo study to assess the capability to recover the true coefficients, true choice data, true cluster membership, etc. for synthetic data we create whose structure is known. Consistent with past methodological testing in the mathematical psychology and psychometric literature (e.g., DeSarbo, 1982; DeSarbo and Carroll, 1985; DeSarbo and Cron, 1988; Jedidi and DeSarbo, 1991) involving Monte Carlo testing of new proposed methods, we created a collection of synthetic datasets as generated by use of an orthogonal factorial design by manipulating independent factors reflecting different data, parameter, and model specification conditions. We employ a 3^7 fractional factorial design (SPSS ORTHOPLAN command; Adelman, 1962) to study the main effects of each factor. Here, we used seven experimentally manipulated factors (three levels per each factor): the number of subjects, the number of segments, the number of stimuli per subject, the number of independent variables, the proportion of active independent variables, the specification of the correlation matrices, and the nature of the hyperparameters of variable selection prior w (see Eq. (6)). As dependent measures, we used some six measures: recovery of choices, the percentage of correctly selected independent variables, the percentage of wrongly selected independent variables, the root mean squared error between the true and recovered cluster level coefficients, cluster membership hit rate, and CPU computing time.

First, we found that the proposed model can successfully recover the true parameters and true data/choices. More specifically, the proposed model provides good performance overall regarding the recovery of data/choices (average of 96.3%), proportion of correctly selected independent variables (average of 94.1%), proportion of incorrectly selected independent variables (average of 94.5%), cluster membership hit rate (average of 94.3%), and RMSE of true coefficient recovery averaging 1.022 across all factors. Second, we checked for methodology robustness by running multiple linear regression analyses (akin to ANOVA) on each of the six dependent measures after coding the seven independent factors as dummy variables to explore their main effects. We found that none of the independent factors were significant at 0.05 level for the first four dependent measures. For the fifth dependent measure (i.e., the cluster membership hit rate), the number of clusters was the only significant factor in the linear regression analysis. As expected, when the number of clusters increases, the membership hit rate to identify true membership is decreased, although the lowest hit rate for $K = 4$ is still quite good (91%). For the sixth dependent measure (i.e., CPU time), the number of subjects was the only significant factor. This is also to be expected as increasing sample size does typically increase computing time. It is interesting to note that none of the levels tested for prior specification of the hyperparameters a and b has significantly affected any of the dependent measures! Overall, this Monte Carlo simulation study demonstrates that the proposed new Bayesian methodology appears somewhat robust with respect to these seven factors and their levels. Appendix B provides more detailed discussion of the study design and specific results.

5. A consumer psychology application

5.1. Study background and data description

We present an automotive consumer research study sponsored by a large US automotive manufacturer to gauge automotive marketing awareness and shopping behavior among typical consumers in the market place. The surveys used in these tracking studies were conducted among new vehicle *intenders* and were nationally collected from an automotive consumer household panel. This "Brand Image Study" had been conducted among new vehicle intenders semiannually in June and December for the past 30 years. An "intender" is defined by the client firm as a consumer that will be "in-market" or has plans to purchase a new vehicle within the next 6–12 months. Each survey respondent rated every brand/model corresponding to the particular product segment in which s/he intended consider purchasing. The ending completed sample resulted in approximately 200–300 respondents per product segment per wave. The information collected includes familiarity with each make/model, advertising recall, overall opinion, purchase consideration, image attribute ratings, and various demographic information.

The Image Study was administered across 16 car segments and 10 light truck and sport utility vehicle (SUV) segments. The Large Sports Utility product segment was selected for use in this study. The data were collected in December 2002 with 290 consumer intenders evaluating 16 different large utility vehicles on the market for sale at that time. Here, the respondents indicated if they would consider each brand in this product segment given they would soon be in-market for purchase. In addition to this consideration to buy the vehicle, the respondents subjectively rated each vehicle on 24 image attributes using 5-point interval measurement scales. DeSarbo, Park, and Rao (2011) had utilized aspects of this data for the Large SUV market to construct positioning maps. See Table 1 for more details of the 24 attributes and the 16 large SUV brands included in the analysis. Numerous demographics were also collected regarding the respondents including gender, marital status, age, income level, occupation, size of household, life cycle status, internet usage, race, geographic location, etc.

The client firm allowed us only one year of data for our research. The consideration to buy data were characterized as "pick any/M" binary choice data reflecting their higher likelihood of further consideration. The goal in this analysis was to derive clusters (here called *market segments*) of intender consumers in terms of the different drivers of their choice decisions with respect to the different SUV attributes deemed important (i.e., integrate multivariate choice (probit) regression, variable selection, and clustering in a simultaneous fashion). In the analyses to follow, we first compare the results of our proposed method with several benchmark procedures (e.g., aggregate logistic regression, aggregate stepwise logistic regression, Bayesian Multivariate Probit regression, Bayesian Multivariate Probit Mixture model *without* Variable Selection, and a two-step procedure using factor analysis and the Bayesian Multivariate Probit Mixture model) using all 16 brands. Then, to compare predictive validation performance, we utilize leave-one-out cross validation.

Table 1
24 attributes and 16 large SUV brands in the large SUV application study.

24 attributes	16 large SUV brands
foreign (Foreign Brand),	Chevy Suburban,
gasmile (Gas Mileage),	Chevy Tahoe,
value4money (Value for the Money),	Ford Expedition,
workmanship (Workmanship),	Range Rover,
ride-handling (Ride/Handling—Streets& Highways),	Toyota Land Cruiser,
load–unload (Easiness to Load & Unload),	Yukon Denali,
luxurious (Luxurious),	Ford Excursion,
safety (Safety/Protection for Occupants),	Lincoln Navigator,
ride off-road (Ride & Handling Off-Road),	LexusLX470,
rugged (Built Rugged and Tough),	Toyota Sequoia,
towing (Excellent Towing Capability),	Hummer H1,
price (Reasonably Priced),	Escalade EXT,
sporty (Sporty),	GMC Yukon,
good-looking (Good Looking),	Mercedes G Class,
family-use (Vehicle for Family Use),	Hummer H2,
fun2drive (Fun to Drive),	Escalade ESV.
passenger room (Interior Passenger Room),	
enter–exit (Easiness to Enter/Exit),	
dependable (Dependability),	
acceleration (Acceleration),	
cargo space (Cargo Space),	
prestigious (Prestigious),	
tech-adv (Technically Advanced),	
trade-in (Trade-in Value).	

5.2. Aggregate analyses

As *aggregate* (i.e., no clusters/segments) benchmark models, we investigate the results produced by applying aggregate logistic regression, aggregate stepwise logistic regression, and a Bayesian multivariate probit model (rmvpGibbs in the R package bayesm, Rossi, 2017). For the Bayesian multivariate probit model, we performed 20,000 iterations and discarded the first 10,000 as a burn-in period. To check convergence of the MCMC chains, we first inspected the corresponding trace plots which showed consistent results after the burn-in period (sample trace plots are provided in Appendix C). Additionally, we performed diagnostic checks proposed by Gelman and Rubin (1992) and Geweke (1992), which provided additional evidence for convergence after burn-in.[1] Table 2 presents the analysis results of these various aggregate binary choice models. Results from the three methods share some resemblance in terms of selected variables and estimated signs and values of coefficients. For the Bayesian multivariate probit model, besides the intercept term, some ten attributes were significant in explaining the consideration to buy binary choices: foreign,

[1] We have implemented two additional convergence diagnostics: namely, diagnostics proposed by Gelman and Rubin (1992), and Geweke (1992). Following Gelman and Rubin (1992), we ran multiple MCMC iterations with different starting points, and found that the computed $\hat{R}$ values of the significant variables showed convergence (see Gelman et al., 2014): e.g., 1.02 for "foreign", 1.04 for "value4money", 1.01 for "load–unload", 1.02 for "price", 1.06 for "good-looking", and 1.02 for "family-use". Also, we followed Geweke (1992) to compare the means between first 10% and last 50% of the chains and found that all Z-scores of the variables were less than 2, which suggests that the means of the first 10% is not significantly different from the last 50%.

Table 2
Results of aggregate logistic regression, aggregate stepwise logistic regression, and bayesian multivariate probit model (rmvpGibbs in R package bayesm, Rossi (2017)).

	Aggregate Logistic Regression			Aggregate Stepwise Logistic Regression			Bayesian Multivariate Probit Model			
	Beta	S.E.	Sig.	Beta	S.E.	Sig.	Post. Mean Beta	Post. S.D.	2.5%	97.5%
intercept	**-5.09**	**0.22**	**0.00**	**-5.04**	**0.21**	**0.00**	**-9.29**	**0.82**	**-11.23**	**-7.64**
foreign	**-0.82**	**0.09**	**0.00**	**-0.83**	**0.09**	**0.00**	**-0.78**	**0.18**	**-1.22**	**-0.41**
gasmile	**-0.04**	**0.02**	**0.03**	0	0	0	-0.04	0.03	-0.11	0.04
value4money	**0.18**	**0.03**	**0.00**	**0.20**	**0.02**	**0.00**	**0.22**	**0.04**	**0.14**	**0.30**
workmanship	0.05	0.03	0.13	0	0	0	0.06	0.04	-0.03	0.15
ride-handling	0.04	0.03	0.20	0	0	0	0.04	0.04	-0.04	0.13
load-unload	0.06	0.03	0.06	**0.08**	**0.03**	**0.01**	**0.11**	**0.04**	**0.02**	**0.20**
lurxurious	0.00	0.03	0.97	0	0	0	0.07	0.04	0.00	0.14
safety	**-0.10**	**0.03**	**0.00**	**-0.07**	**0.03**	**0.02**	**-0.10**	**0.05**	**-0.20**	**-0.01**
ride-offroad	-0.05	0.03	0.07	**-0.05**	**0.02**	**0.03**	-0.05	0.04	-0.13	0.03
rugged	0.02	0.03	0.44	0	0	0	0.07	0.04	-0.02	0.16
towing	0.05	0.03	0.06	**0.07**	**0.03**	**0.01**	**0.12**	**0.04**	**0.03**	**0.21**
price	**0.05**	**0.02**	**0.01**	0	0	0	**0.08**	**0.03**	**0.01**	**0.14**
sporty	-0.03	0.02	0.27	0	0	0	-0.01	0.03	-0.08	0.06
good-looking	**0.27**	**0.03**	**0.00**	**0.26**	**0.02**	**0.00**	**0.31**	**0.04**	**0.22**	**0.41**
family-use	**0.10**	**0.03**	**0.00**	**0.11**	**0.03**	**0.00**	**0.15**	**0.04**	**0.07**	**0.23**
fun2drive	-0.01	0.03	0.66	0	0	0	0.05	0.04	-0.03	0.14
passenger room	0.02	0.04	0.66	0	0	0	-0.02	0.05	-0.12	0.08
enter-exit	**-0.11**	**0.03**	**0.00**	**-0.09**	**0.03**	**0.00**	**-0.20**	**0.05**	**-0.31**	**-0.11**
dependable	0.06	0.03	0.10	0	0	0	0.06	0.05	-0.04	0.16
acceleration	**0.09**	**0.03**	**0.00**	**0.09**	**0.03**	**0.00**	**0.18**	**0.05**	**0.08**	**0.28**
cargo space	0.04	0.03	0.19	**0.06**	**0.03**	**0.03**	0.07	0.04	-0.02	0.17
prestigious	-0.03	0.03	0.29	0	0	0	-0.01	0.04	-0.10	0.07
tech-adv	0.04	0.03	0.22	0	0	0	0.05	0.05	-0.05	0.15
trade-in	-0.05	0.03	0.12	0	0	0	0.01	0.04	-0.07	0.10

Note: Shaded and Bold cells denote significance at $p < 0.05$ for both Logistic Regression and Stepwise Logistic Regression; For the Bayesian Multivariate Probit Model, Shaded and Bold cells do not include zero values in the 95% credible intervals.

value4money, load–unload, safety, towing, price, good-looking, family-use, enter–exit, and acceleration. That is, for this Bayesian multivariate probit solution, non-foreign brands that are perceived to be 'good value for the money', 'easy to load and unload', 'excellent in towing capability', 'reasonably priced', 'good looking', 'good for family use', and 'excellent in acceleration' appear to be significant for more serious consideration for purchase among this aggregate sample of intenders. However, we observe two negative coefficient estimates for safety (good safety) and enter–exit (easy to enter/exit), respectively, which are somewhat counter intuitive concerning face validity interpretation. The aggregate logistic regression solution is somewhat similar with some exceptions. Here, three negative coefficient estimates (gasmile, safety, and enter–exit) are counter intuitive. Lastly, the aggregate stepwise logistic regression solution is similar to the above two solutions, and again it is difficult to interpret the three significant negative coefficients for safety, ride off-road, and enter–exit.

5.3. Analysis by the Bayesian multivariate probit mixture model without variable selection

In order to investigate the usefulness of simultaneous variable selection for the proposed model, we now compare our disaggregate Bayesian multivariate probit mixture model *without* variable selection to our proposed disaggregate model *with* variable selection. We operated the *without* variable selection model in our software by setting all Z's, the variable selection indicators in the proposed model, to 1 (i.e., no variables are excluded). Given that the $K = 2$ solution of the proposed model with variable selection will be selected (see Table 5), the $K = 2$ solution of the disaggregate multivariate probit mixture model *without* variable selection is reported in Table 3 for comparison purposes, and the posterior standard deviations and 95% credible intervals are used to determine active variables. For this mixture model, we ran our MCMC with 20,000 iterations and discarded the first 10,000 as a burn-in period—the same as for our proposed model, and we checked convergence in the same methods as the proposed model.

From Table 3, Segment/Cluster 1 consists of 38% of the total sample, and here six variables (besides the intercept) are significant (foreign, value4money, ride-offroad, rugged, good-looking, and family-use). Members of this segment positively consider domestic brands that are perceived as good in value for the money, being built rugged and tough, good-looking, and good for family use. But, it is somewhat questionable to find the large negative impact of good ride off-road on choice probability. The Segment/Cluster 2 consists of 62% of the sample, and eleven variables besides the intercept (foreign, value4money, workmanship, safety, towing, price, good-looking, family-use, enter–exit, dependable and acceleration) are significant. This segment also values the non-foreign brands that are good in value for the money, excellent in workmanship, excellent towing capability, reasonably priced, good-looking, good for family use, and dependable, excellent in acceleration. However, we still observe counter intuitive

Table 3
$K = 2$ solution by Bayesian multivariate probit mixture model *without* variable selection (all Z's are set to 1 in the proposed model).

	Segment 1				Segment 2			
	Coeffs	Post. S.D.	2.50%	97.50%	Coeffs	Post. S.D.	2.50%	97.50%
intercept	**-5.65**	**0.45**	**-6.59**	**-4.91**	**-3.61**	**0.18**	**-3.97**	**-3.29**
foreign	**-0.86**	**0.20**	**-1.30**	**-0.53**	**-0.34**	**0.08**	**-0.49**	**-0.18**
gasmile	0.00	0.03	-0.07	0.06	-0.02	0.02	-0.05	0.02
value4money	**0.11**	**0.04**	**0.02**	**0.17**	**0.13**	**0.03**	**0.08**	**0.18**
workmanship	-0.03	0.06	-0.13	0.06	**0.06**	**0.03**	**0.00**	**0.11**
ride-handling	0.06	0.04	-0.05	0.13	0.01	0.03	-0.05	0.08
load-unload	0.05	0.04	-0.03	0.12	0.04	0.03	-0.02	0.10
luxurious	-0.01	0.03	-0.07	0.06	0.03	0.02	-0.02	0.07
safety	0.02	0.05	-0.09	0.11	**-0.09**	**0.03**	**-0.15**	**-0.04**
ride-offroad	**-0.09**	**0.04**	**-0.16**	**-0.03**	-0.03	0.03	-0.08	0.04
rugged	**0.11**	**0.04**	**0.03**	**0.21**	0.02	0.03	-0.03	0.07
towing	0.08	0.05	-0.01	0.20	**0.06**	**0.02**	**0.01**	**0.10**
price	-0.02	0.03	-0.10	0.04	**0.06**	**0.02**	**0.02**	**0.10**
sporty	0.00	0.04	-0.08	0.08	-0.01	0.02	-0.06	0.03
good-looking	**0.16**	**0.05**	**0.08**	**0.25**	**0.19**	**0.03**	**0.12**	**0.24**
family-use	**0.18**	**0.04**	**0.09**	**0.26**	**0.06**	**0.02**	**0.02**	**0.11**
fun2drive	0.02	0.03	-0.04	0.09	0.00	0.02	-0.03	0.05
passenger room	0.04	0.04	-0.03	0.11	-0.02	0.03	-0.07	0.05
enter-exit	-0.06	0.05	-0.14	0.06	**-0.10**	**0.04**	**-0.18**	**-0.01**
dependable	-0.01	0.05	-0.10	0.10	**0.06**	**0.03**	**0.00**	**0.11**
acceleration	-0.02	0.04	-0.10	0.06	**0.09**	**0.03**	**0.03**	**0.15**
cargo space	0.03	0.04	-0.05	0.11	0.03	0.03	-0.03	0.07
prestigious	0.00	0.04	-0.07	0.10	-0.03	0.03	-0.09	0.03
tech-adv	0.05	0.05	-0.03	0.15	-0.01	0.04	-0.08	0.06
trade-in	-0.04	0.06	-0.18	0.08	0.00	0.03	-0.05	0.06
Mixture	38%				62%			

Note: Shaded and Bold cells do not include zero values in 95% credible intervals.

negative coefficients with respect to safety and easiness of enter/exit, which are problematic for interpretation/implementation. These questionable negative coefficients were consistently found across the aggregate analyses reported in Table 2, and these negative coefficients are split into two segments. We believe that including irrelevant or correlated variables might affect such questionable coefficient values and variable selection can help to resolve this issue.

5.4. A two-step procedure using factor analysis and the Bayesian multivariate probit mixture model

Another possible approach to the problem of variable selection and collinearity involves applying factor /principal components analysis to the set of independent variables. There are some obvious limitations to such an approach. First, one has to use a two-step approach (i.e., factor analysis first and then submit the estimated factor scores into the choice analysis model), while our proposed methodology *simultaneously* selects relevant independent variables and conducts the multivariable probit mixture analysis. As such, the factor analysis itself is conducted independently without any relation to the dependent variables. Second, when using dimension reduction approaches such as factor analysis, users need to subjectively decide on what form of factor analysis is most appropriate as well as what form of rotation procedure to utilize. Not only does this affect the labeling of the underlying abstract dimensions, but one can lose specific attribute information which can limit interpretability and (in this application) the managerial actionablility of the model outputs. However, given that factor analysis is commonly recognized as a potential solution to handle many inter-correlated independent variables, we applied factor analysis before running the Bayesian multivariate probit mixture model *without* variable selection. For this factor analysis, we used the maximum likelihood approach with Varimax rotation in SPSS, and selected a four-factor model explaining total 68% of data variation, based on the conventional rule "eigenvalue $\geq$ 1". Here, Factor 1 is the most important factor explaining the largest (49%) amount of total variance. Using top three attribute loadings of prestigious, tech_adv, and trade_in, we label this factor *Expensive*. In a similar fashion, we name Factor 2 as *Convenience* (top three contributing attributes: family use, passenger_room, and cargo_space), Factor 3 as *Value* (top three contributing attributes: gasmile, value4money, and price), and Factor 4 as *Toughness* (top three contributing attributes: ride_offroad, rugged, and towing).

We submitted the four columns of factor scores as independent variables (with an intercept) and multivariate choices as the dependent variable to the multivariate probit mixture model *without* variable selection. Given that the $K = 2$ solution of the proposed model with variable selection will be selected (see Table 5), the $K = 2$ solution of the disaggregate multivariate probit mixture model *without* variable selection is reported in Table 4 for comparison purposes, and that posterior standard deviations and 95% credible intervals are used to determine active variables. For this mixture model analysis, we ran the MCMC with 20,000 iterations and discarded the first 10,000 as a burn-in period—the same as for our proposed model, and we checked convergence using the

Table 4
$K = 2$ solution via the two-step approach: Step 1. Factor analysis, Step 2. Bayesian multivariate probit mixture model without variable selection.

	Segment 1				Segment 2			
	Coeffs	Post. S.D.	2.50%	97.50%	Coeffs	Post. S.D.	2.50%	97.50%
Intercept	−0.950	0.054	−1.070	−0.855	0.275	0.136	−0.012	0.538
Factor1	0.342	0.039	0.267	0.417	0.520	0.084	0.353	0.686
Factor2	0.594	0.040	0.516	0.672	0.491	0.075	0.342	0.636
Factor3	0.436	0.038	0.363	0.514	0.827	0.111	0.628	1.056
Factor4	0.179	0.034	0.114	0.248	0.385	0.074	0.241	0.535
Mixture	78%				22%			

Table 5
DIC and BIC for the proposed Bayesian multivariate probit mixture model.

	K=1	K=2	K=3	K=4	K=5
DIC	2480.1	**2291.7**	2299.8	2339.2	2408.5
BIC	2522.8	**2429.7**	2473.1	2500.7	2682.4

Note: Shaded and bold cells indicate the model selection criteria's suggested solution.

same methods as the proposed model. As presented in Table 4, Segment/Cluster 1 consists of 78% of the total sample and all four factors are significant. Segment/Custer 2 consists of 22% of total sample and all four factors are significant as well. From this analysis, we would conclude that Factor 2 is a more important driver for Segment 1 (0.59 for Segment 1 vs. 0.49 for Segment 2), and Factor 1, Factor 3, and Factor 4 are comparatively more important drivers for Segment 2 (coefficient effect sizes comparisons: 0.34 for Segment 1 vs. 0.52 for Segment 2 for Factor 1; 0.436 for Segment 1 vs. 0.827 for Segment 2 for Factor 3; 0.179 for Segment 1 vs. 0.385 for Segment 2 for Factor 4). While all signs of the coefficients for the four derived factors are positive across the two segments, it is difficult for management to manipulate specific attribute levels differentially across the two derived segments given the difficulty of dealing with abstract factors. Predictive validation ability will be discussed and compared with the proposed model shortly.

5.5. The proposed Bayesian model analysis with variable selection

For our proposed multivariate probit mixture model *with* variable selection, we ran our MCMC with 20,000 iterations and discarded the first 10,000 as a burn-in period, and the hyperparameters were set as: $a = 10, b = 50; s_{p1} = 1, s_{p2} = 1$. As a sparse prior, we expect w to be small and the values of a and b are chosen to reflect such prior belief. Also, the choice ofs_{p1} and s_{p2} was made to represent somewhat vague prior information. To check convergence of the MCMC chains, we first inspected the corresponding trace plots which showed consistent results after the burn-in period (see sample trace plots are provided in Appendix C). Additionally, we computed convergence diagnostics as proposed in Gelman and Rubin (1992) and Geweke (1992), which provided additional evidence for convergence after burn-in.[2] Table 5 presents the model selection criteria for selecting the number of segments (K). As shown, both criteria indicate the $K = 2$ segment solution as the optimal solution for this dataset. Note that the aggregate ($K = 1$) solution is rejected here in favor of this $K = 2$ segment solution.

Table 6 displays the estimated coefficients by derived market segment for all selected attributes. Here, we see two sizable segments: members of Segment/Cluster 1, which contains 59% of the sample, tend to avoid foreign brands and appear to pursue SUV vehicles that are easy for loading & unloading, good looking, and good for family use. Given those selected variables and their coefficients, Segment 1 can be described as a *Family-Oriented* consumer segment pursuing convenience. Consistent with such a family-oriented tendency, members of Segment 1 have a significantly higher marriage rate (80.7% vs. 68.1%, $p < 0.03$), and are more likely to have a higher average household size (3.31 vs. 3.0, $p < 0.07$). In addition, they are less likely to use the Internet (55% vs. 70.6%, $p < 0.05$) and more likely to be Caucasian (83% vs. 76.5%, $p < 0.5$). Thus, firms can effectively communicate to the members of Segment 1 with family-valued messages through traditional communication channels. In contrast, members of Segment 2, comprising 41% of the sample, tend to pursue SUV brands with good value for the money, a reasonable price, and good looking. This segment can be described as a SUV consumer group who pursue *Economic Value*. Members of Segment/Cluster 2 are likely to have lower-marriage rates, have a stronger tendency to use Internet, and possess lower average household sizes. So, the analysis results suggest that firms should focus on a pricing strategy and communicate the economic value of SUV's using more online communication channels for Segment 2. Last, all potential consumers in this dataset appear to care about social impression since the attribute of good looking is significant across both segments. We note the lack of any problematic (negative) coefficient signs, as present in many of the previous benchmark solutions, which aids in the interpretation here.

Table 7 displays the consideration choice proportion for each derived segment for the 16 brands of large SUV's for this $K = 2$ solution derived by the proposed method. As illustrated from Table 7, intenders in Segment 1 have much higher consideration for Chevrolet Suburban, Chevrolet Tahoe, Ford Expedition, Yukon Denali, Ford Excursion, and GMC Yukon—all *domestic* brands. In direct contrast, the major consideration brands set of Segment 2 clearly includes *foreign* brands—Range Rover, Land Cruiser, Lexus LX 470, Toyota Sequoia, and Mercedes G Class.

Rather than having to inspect large upper/lower triangular half matrices of the estimated correlation matrices, Fig. 1(a) presents a two-dimensional market structure representation based on an analysis of the posterior means of the correlation matrix for Segment 1 by using the ALSCAL procedure (i.e., one of the multidimensional scaling methods) in SPSS. Fig. 1(b) depicts the analogous spatial structure for Segment 2. The two figures reveal the underlying structure of the estimated correlations of the latent

[2] Following Gelman and Rubin (1992), we found that the computed $\hat{R}$ values of the significant variables of the proposed model showed a reasonably good convergence (see Gelman et al., 2014): e.g., for $\hat{R}$ values of Segment1, 1.17 for "foreign", 1.34 for "load–unload", 1.03 for "good-looking", and 1.18 for "family-use"; for Segment 2, 1.15 for "value4money", 1.09 for "price", and 1.13 for "good-looking". In addition, we followed Geweke (1992) to compare the means between first 10% and last 50% of the chains and found that all Z-scores of the variables were less than 2. Note, similar results of convergence were found in the Bayesian multivariate probit mixture model *without* variable selection which was used as a comparative benchmark.

Table 6
$K = 2$ results by the proposed Bayesian multivariate probit mixture model.

	Segment 1			Segment 2		
	Coeffs	Post. S.D.	Odds Ratio	Coeffs	Post. S.D.	Odds Ratio
intercept	**-4.16**	**0.28**	**>1000**	**-3.41**	**0.26**	**>1000**
foreign	**-0.67**	**0.25**	**550.56**	0.00	0.00	0.00
gasmile	0.00	0.00	0.00	0.00	0.00	0.00
value4money	0.00	0.00	0.00	**0.21**	**0.04**	**126.58**
workmanship	0.00	0.00	0.00	0.00	0.00	0.00
ride-handling	0.00	0.00	0.00	0.00	0.00	0.00
load-unload	**0.14**	**0.05**	**127.74**	0.00	0.00	0.00
luxurious	0.00	0.00	0.00	0.00	0.00	0.00
safety	0.00	0.00	0.00	0.00	0.00	0.00
ride-offroad	0.00	0.00	0.00	0.00	0.00	0.00
rugged	0.00	0.00	0.00	0.00	0.00	0.00
towing	0.00	0.00	0.00	0.00	0.00	0.00
price	0.00	0.00	0.00	**0.10**	**0.02**	**132.61**
sporty	0.00	0.00	0.00	0.00	0.00	0.00
good-looking	**0.18**	**0.03**	**>1000**	**0.19**	**0.03**	**>1000**
family-use	**0.11**	**0.04**	**30.97**	0.00	0.00	0.00
fun2drive	0.00	0.00	0.00	0.00	0.00	0.00
passenger room	0.00	0.00	0.00	0.00	0.00	0.00
enter-exit	0.00	0.00	0.00	0.00	0.00	0.00
dependable	0.00	0.00	0.00	0.00	0.00	0.00
acceleration	0.00	0.00	0.00	0.00	0.00	0.00
cargo space	0.00	0.00	0.00	0.00	0.00	0.00
prestigious	0.00	0.00	0.00	0.00	0.00	0.00
tech-adv	0.00	0.00	0.00	0.00	0.00	0.00
trade-in	0.00	0.00	0.00	0.00	0.00	0.00
Mixture	59%			41%		

Note: Shaded and Bold cells denote odds ratio > 20.

Table 7
Choice proportions of derived segments for the $K = 2$ solution by the proposed model.

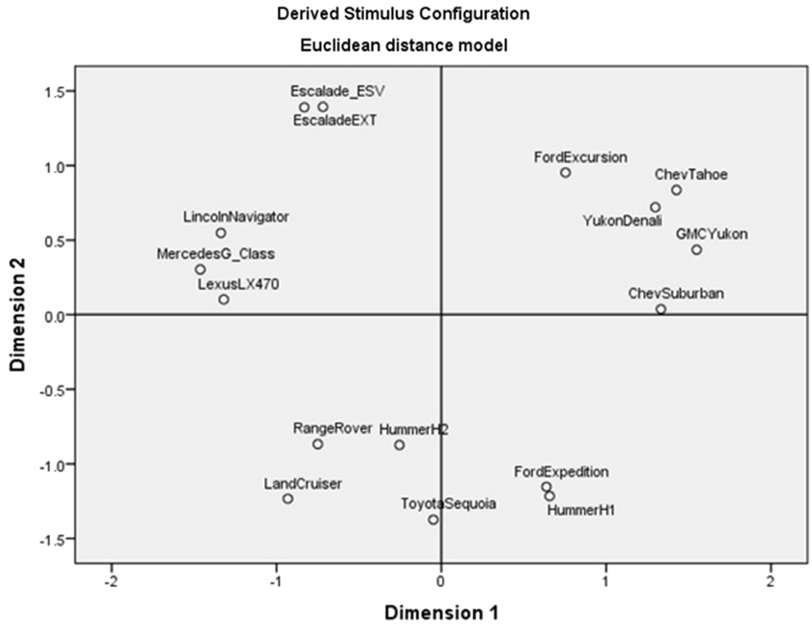
Brand	Segment 1	Segment 2
ChevSuburban	64%	36%
ChevTahoe	67%	34%
FordExpedition	62%	38%
RangeRover	***16%***	***84%***
LandCruiser	***33%***	***67%***
YukonDenali	62%	38%
FordExcursion	65%	35%
LincolnNavigator	51%	50%
LexusLX470	***21%***	***80%***
ToyotaSequoia	***36%***	***64%***
HummerH1	52%	48%
EscaladeEXT	52%	49%
GMCYukon	66%	34%
MercedesG_Class	***22%***	***78%***
HummerH2	48%	52%
Escalade_ESV	50%	50%

Note: Bold font indicates foreign brands.

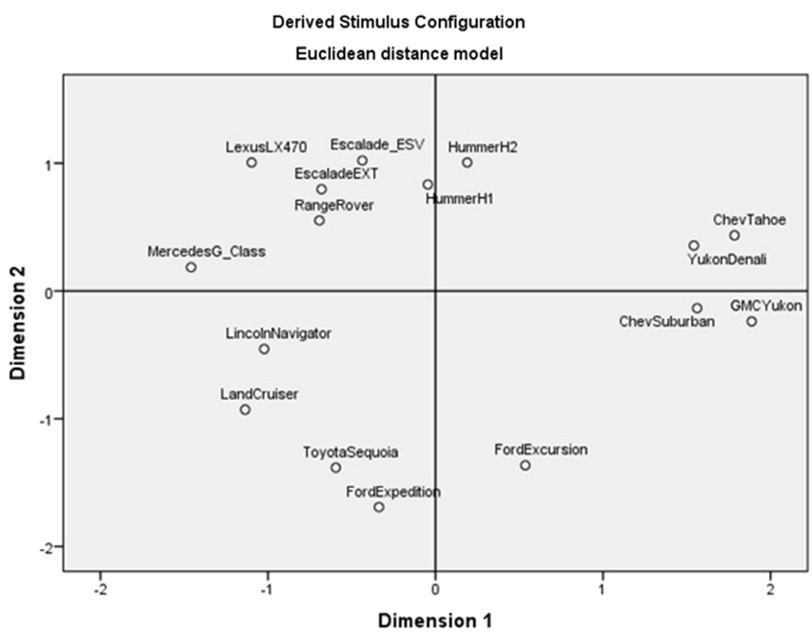
errors in Eq. (2). Here, brands that are located close to one another appear to display similar choice patterns amongst these consumers, and would therefore be assumed to be more competitive with one another. Dimension I in both figures appears to mostly separate the foreign brands (left) from the domestic brands (right) with some exceptions. Consistent with the consideration choice proportions from Table 7, most of the domestic brands which are highly considered by Segment 1 (e.g., Chevrolet Suburban, Chevrolet Tahoe, Yukon Denali, GMC Yukon, and Ford Excursion) are closely located in or near Quadrant I of the map in Fig. 1(a). In contrast, all of foreign brands which are highly considered by Segment 2 (e.g., Range Rover, Lexus LX 470, Land Cruiser, Mercedes G Class, and Toyota Sequoia) are spread across Quadrants II and III of the map in Fig. 1(b), suggesting more differentiation among these foreign brands. The bad news for GM here is that their brands appear to compete mostly with each other and cannibalize market share in both figures suggesting the need for an improved positioning strategy involving the need for better differentiation of these offerings.

Finally, let's compare and discuss in-sample model fit and out-of-sample (leave-one-out) predictive validation performances between the proposed model and the other benchmark procedures. First, regarding the in-sample model fit, we used the in-sample correct hit rate (pseudo R^2; Freese and Long, 2006). As presented in the first row of Table 8, the three aggregate probit models provide somewhat inferior model fits compared to three finite mixture models: 75% for aggregate logistic regression, 74.8% for aggregate stepwise logistic regression, 74.6% for aggregate Bayes multivariate probit regression vs. 78% for the multivariate probit mixture model *without* variable selection, 76.7% for the two-step procedure with factor analysis first and multivariate probit mixture model, and 78.2% for the proposed model. Here, although the proposed model provides the best in-sample model fit, it is observed that the multivariate probit mixture model *without* variable selection provides a very similar in-sample fit.

Next, let's compare out-of-sample (leave-one-out) predictive validation results. For the predictive validation of the first two benchmark models (i.e., logistic regression, stepwise logistic regression), we compute the choice probabilities by using the equation of $\frac{exp\left(\underline{X}_i\hat{\underline{\beta}}\right)}{1+exp\left(\underline{X}_i\hat{\underline{\beta}}\right)}$ (here, $\underline{X}_i$ is the vector of covariates for a validation brand), and if the probability is greater than 0.5, the



(a) Based on estimated posterior correlation matrix for segment 1.



(b) Based on estimated posterior correlation matrix for segment 2.

Fig. 1. Two-dimensional mapping using ALSCAL in SPSS.

predicted choice is 1, otherwise 0. For the proposed model and the multivariate probit mixture model *without* variable selection, we assume that the latent utility $\underline{Y}_{i*}$ follows a mixture distribution as in Eq. (2) – but for a single component only – and the choice response c_{i*} is defined as in Eq. (1). Then, we generate $\underline{Y}_{i*}$from its predictive distribution, evaluate the corresponding c_{i*} for each generated value of $\underline{Yi*}$, and compute the relative frequency of $c_{i*} = 1$. Furthermore, the value of 0.5 is used as the cutoff point between choice and non-choice. In the case of the (aggregate) Bayesian multivariate probit model, the prediction procedure is the same except that there is no need to consider the mixture distribution. The simple hit ratio and the asymmetric Jaccard similarity index are utilized as the measures of predictive validation as researchers are typically more interested in predicting choice states (i.e., $c_{i*} = 1$). The Jaccard index is calculated by the formula: $\frac{J_{11}}{J_{01}+J_{10}+J_{11}}$, where J_{11} is total number of correct cases when the model predicted a choice ($\hat{c}_{im} = 1$) and the data also recorded a choice ($c_{im} = 1$); J_{10} is total number of incorrect cases when the model predicted a choice ($\hat{c}_{im} = 1$) but the data recorded a non-choice ($c_{im} = 0$); and J_{01} is total number of incorrect cases when the model predicted a non-choice ($\hat{c}_{im} = 0$) but the data recorded a choice ($c_{im} = 1$).

As shown in the second and third rows of Table 8, the proposed model with variable selection provides a higher leave-one-out predictive validation performance in being able to accurately predict 78.1% of hit rate and 36% of the choice consideration predictability (Jaccard similarity index). Here, the aggregate logistic regression, aggregate stepwise logistic regression and Bayesian multivariate probit model show lower 74.3%, 74.1%, and 74% hit rates, and

Table 8
In-sample model fit and out-of-sample (leave-one-out) choice predictive validation performance comparison between the proposed model and Benchmark models.

	Aggregate logistic regression	Aggregate stepwise logistic regression	Bayesian multivariate probit regression (Rossi, 2017)	Multivariate probit mixture model (without variable selection)	Two-step: factor analysis & multivariate probit mixture	The proposed model
In-sample choice hit rate: Pseudo R^2	75%	74.8%	74.6%	78%	76.7%	78.2%[a]
Leave-one-out choice hit rate	74.3%	74.1%	74%	74.9%	74.6%	78.1%[a]
Leave-one-out Jaccard similarity index	27.6%	27.3%	28%	30.9%	30.4%	36%[a]

Note:
[a] Indicates best performance value amongst the five competing models.

27.6%, 27.3%, and 28% Jaccard similarity indices, respectively. The disaggregate multivariate probit mixture model *without* variable selection also has a lower 74.9% hit rate and 30.9% Jaccard similarity index. As a possible explanation for the superior performance of the proposed model, we found severe (multi)collinearity among the 24 attributes (all significant pairwise correlations with $p < 0.01$; many condition indices are greater than 30) and so segment-level variable selection may contribute positively to the prediction performance (as well as the face validity of the results). To address this collinearity issue, one might think that dimension reduction (e.g., factor analysis) can be a possible alternative to variable selection approach, but such two-step procedure using factor analysis before mixture probit analysis provides a lower 74.6% hit rate and 30.4% Jaccard similarity index compared to the proposed model. As previously discussed, the factor analysis needed to be conducted separately without considering the relationship with the dependent variable (consumer choices), and this stagewise usage presents potential difficulties regarding interpretability and actionability for practitioners as most derived factors tend to be somewhat abstract compared to the utilization of specific attributes. Thus, it is more difficult to know how to alter linear combinations of attributes fol repositioning existing brands and/or positioning new brands.

6. Discussion and conclusions

We propose a new heterogeneous Bayesian finite mixture model conforming to the information processing described in the current behavioral decision literature in psychology for multiple correlated choice data with large numbers of features/attributes. The goal of this new model is to identify a subset of key predictors and estimates of the coefficients for each probit regression model per derived cluster while accommodating the dependency structure across multiple choices per each derived cluster—*all simultaneously*. In addition, we illustrate how to conveniently derive market structure from the estimated posterior means of the correlation matrices through the spatial representation of such matrices. Concerning this SUV consumer psychology application, the proposed model practically can help firms to develop efficient marketing resource allocation by emphasizing segment-wise sparse key drivers in its promotions and improve positioning strategy (e.g., combining complementary choices, differentiating cannibalized choices within manufacturer, etc.) by providing dependency structures across choices (brands) by market segment.

We presented a Monte Carlo simulation study to demonstrate that the proposed model is able to successfully recover true parameters/true choices and identify true segment membership in a somewhat robust manner. Then, we have provided an actual consumer psychology application with respect to the consideration choices of actual intenders of purchasing large SUV brands, and our proposed model has shown better performance over several benchmark models including the aggregate logistic regression, aggregate stepwise logistic regression, Bayesian multivariate probit regression, the disaggregate Bayesian multivariate probit mixture model *without* variable selection, and the two-step procedure with factor analysis first and multivariate probit mixture analysis second. We provide substantive insights into the heterogeneous nature of consideration choice set formation for the SUV market place, as well as favorable comparative prediction validation performances in comparison to these various benchmark models. The proposed model should become a useful tool for the analysis of correlated choice data which are commonly collected in a variety of social science applications.

Appendix A. Derivations of the full conditional distributions

(1) Full conditional distribution of Y_i:

$$\pi\left(\underline{Y}_i|\underline{c}_i \textit{ and others}\right) \propto \pi(\underline{Y}_i|\underline{\beta}_{G_i}, \underline{R}_{G_i}) \times \prod_{m=1}^{M}\left[1_{(c_{im}=1)}1_{(Y_{im}>0)} + 1_{(c_{im}=0)}1_{(Y_{im}\le 0)}\right] \times \propto TMVN\left(\underline{X}_i^T\underline{\beta}_{G_i}, \underline{R}_{G_i}\right),$$

where *TMVN* represents truncated multivariate normal distribution. Each component of $\underline{Y}_i$, Y_{im}, is simulated from $Y_{im}|Y_{i[-m]}, \underline{\beta}_{G_i}, \underline{R}_{G_i}$, which is a univariate Gaussian distribution truncated to $(0, \infty)$ if $c_{im} = 1$ and to $(-\infty, 0]$ if $c_{im} = 0$.

(2) Full conditional distribution of G_i:
Since:

$$\pi\left(G_i = k|\textit{ all others}\right) \propto \pi\left(\underline{Y}_i|G_i = k, \underline{\beta}_k, \underline{R}_k\right) \times \pi\left(G_i = k|d_k\right),$$

we have $\pi\left(G_i = k|\textit{ all others}\right) \propto d_k^*$, where:

$$d_k^* = \frac{\exp\left[-\frac{1}{2}\left\{(\underline{Y}_i - \underline{X}_i\underline{\beta}_k)^T\underline{R}_k^{-1}(\underline{Y}_i - \underline{X}_i\underline{\beta}_k)\right\}\right]d_k}{\sum_{l=1}^{K}\exp\left[-\frac{1}{2}\left\{(\underline{Y}_i - \underline{X}_i\underline{\beta}_l)^T\underline{R}_l^{-1}(\underline{Y}_i - \underline{X}_i\underline{\beta}_l)\right\}\right]d_l}.$$

(3) Full conditional distribution of $\underline{d}$:

$$\pi\left(\underline{d}|\textit{ all others}\right) \propto \pi(\underline{G}|\underline{d})\pi(\underline{d}) \propto \prod_{k=1}^{K} d_k^{\alpha_k+n_k-1}.$$

where $n_k = \sum_{i=1}^{n}\mathbf{1}_{\{G_i=k\}}$. Thus,

$$\pi\left(\underline{d}|\textit{ all others}\right) = \textit{Dirichlet}\left(\alpha_1 + n_1, \ldots, \alpha_K + n_K\right).$$

(4) Full conditional distribution of β_{kp}:
If $Z_{kp} = 0$,

$$\pi\left(\beta_{kp} = 0|Z_{kp} = 0 \textit{ and others}\right) = 1,$$

If $Z_{kp} = 1$,

$$\pi\left(\beta_{kp}|Z_{kp} = 1 \textit{ and others}\right) \propto \phi(\beta_{kp}|0, \tau_p^2)\prod_{i\epsilon id_k}\phi(\underline{Y}_i|\underline{X}_i\underline{\beta}_k, \underline{R}_k) \propto exp\left\{-\frac{\beta_{kp}^2}{2\tau_p^2}\right\}$$

$$\times\ exp\left\{\sum_{i\epsilon id_k} -\frac{1}{2}(\underline{Y}_i - \underline{X}_{i[-p]}\underline{\beta}_{k[-p]} - \underline{X}_{ip}\beta_{kp})^T \right.$$

$$\left. \times\ \underline{R}_k^{-1}(\underline{Y}_i - \underline{X}_{i[-p]}\underline{\beta}_{k[-p]} - \underline{X}_{ip}\beta_{kp})\right\}$$

$$\propto exp\left\{-\frac{\beta_{kp}^2}{2\tau_p^2}\right\} exp\left\{-\frac{1}{2}tr\left[\underline{R}_k^{-1}\sum_{i\epsilon id_k}\left\{\underline{X}_{ip}\underline{X}_{ip}^T\beta_{kp}^2\right.\right.\right.$$

$$\left.\left.\left. -\ 2(\underline{Y}_i - \underline{X}_{i[-p]}\underline{\beta}_{k[-p]})X_{ip}^T\beta_{kp}\right\}\right]\right\}$$

$$\propto exp\left\{-\frac{1}{2}\left[\left(\frac{1}{\tau_p^2}+\sum_{i\epsilon id_k}\underline{X}_{ip}^T\underline{R}_k^{-1}\underline{X}_{ip}\right)\beta_{kp}^2\right.\right.$$

$$\left.\left. -\ 2\left(\sum_{i\epsilon id_k}X_{ip}^T\underline{R}_k^{-1}(\underline{Y}_i - \underline{X}_{i[-p]}\underline{\beta}_{k[-p]})\right)\beta_{kp}\right]\right\}$$

$$\propto exp\left\{-\frac{1}{2}\left(\beta_{kp} - \frac{\sum_{i\epsilon id_k}\underline{X}_{ip}^T\underline{R}_k^{-1}(\underline{Y}_i - \underline{X}_{i[-p]}\underline{\beta}_{k[-p]})}{\frac{1}{\tau_p^2}+\sum_{i\epsilon id_k}\underline{X}_{ip}^T\underline{R}_k^{-1}\underline{X}_{ip}}\right)^2\right.$$

$$\left. /\left(\frac{1}{\tau_p^2}+\sum_{i\epsilon id_k}\underline{X}_{ip}^T\underline{R}_k^{-1}\underline{X}_{ip}\right)^{-1}\right\},$$

where $id_k = \{i : G_i = k\}$, $\underline{X}_{i[-p]}^T$ is $\underline{X}_i^T$ without the pth column vector, and $\underline{\beta}_{k[-p]}$ is $\underline{\beta}_k$ without its pth component. Thus,

$$\pi\left(\beta_{kp}|Z_{kp}=1\ and\ others\right)$$
$$= N\left(\frac{\sum_{i\epsilon id_k}\underline{X}_{ip}^T\underline{R}_k^{-1}(\underline{Y}_i - \underline{X}_{i[-p]}\underline{\beta}_{k[-p]})}{\frac{1}{\tau_p^2}+\sum_{i\epsilon id_k}\underline{X}_{ip}^T\underline{R}_k^{-1}\underline{X}_{ip}},\right.$$
$$\left.\left(\frac{1}{\tau_p^2}+\sum_{i\epsilon id_k}\underline{X}_{ip}^T\underline{R}_k^{-1}\underline{X}_{ip}\right)^{-1}\right).$$

(5) Conditional distribution of Z_{kp}:
Since

$$\pi\left(Z_{kp},\beta_{kp}|\ all\ others\right)$$
$$= \pi\left(\beta_{kp}|\ all\ others\ including\ Z_{kp}\right)$$
$$\times\ \pi\left(Z_{kp}|\ all\ others\ except\ \beta_{kp}\right)$$

and β_{kp} is generated as described in (4) above, we only need to generate Z_{kp} from $\pi\left(Z_{kp}|\ all\ others\ except\ \beta_{kp}\right)$. Because Z_{kp} can take on two possible values, 0 and 1 only, so $\pi\left(Z_{kp}|\ all\ others\ except\ \beta_{kp}\right) = Bernoulli\left(\frac{1}{1+Q}\right)$ where:

$$Q = \frac{\pi\left(Z_{kp}=0|\ all\ others\ except\ \beta_{kp}\right)}{\pi\left(Z_{kp}=1|\ all\ others\ except\ \beta_{kp}\right)}.$$

To compute Q, note that $\pi\left(Z_{kp}|\ all\ others\ except\ \beta_{kp}\right) = \pi\left(\beta_{kp},Z_{kp}|\ all\ others\right)/\pi(\beta_{kp}|\ all\ others\ including\ Z_{kp})$, and:

$$\pi\left(\beta_{kp},Z_{kp}|\ all\ others\right)$$
$$\propto \prod_{i\epsilon id_k}\phi\left(\underline{Y}_i|\underline{X}_i^T\underline{\beta}_k,\underline{R}_k\right)\times[Z_{kp}\phi\left(\beta_{kp}|0,\tau_p^2\right)$$
$$+\ \left(1-Z_{kp}\right)\mathbf{1}_{\{\beta_{kp}=0\}}]\times w^{Z_{kp}}\times(1-w)^{(1-Z_{kp})}.$$

Thus,

$$\log(Q) = \log\pi\left(Z_{kp}=0|\ all\ others\ except\ \beta_{kp}\right)$$
$$-\ \log\pi\left(Z_{kp}=1|\ all\ others\ except\ \beta_{kp}\right)$$
$$= \log(1-w) - \log(w)$$
$$-\ \frac{1}{2}\left(\log\left|\underline{R}_{n_k}\right| - \log\left|\underline{X}_{id_k[p]}\underline{X}_{id_k[p]}^T\tau_p^2+\underline{R}_{n_k}\right|\right)$$
$$-\ \frac{1}{2}\left(\underline{Y}_{[id_k]} - \underline{X}_{id_k[-p]}\underline{\beta}_{k[-p]}\right)^T$$
$$\times\left[\underline{R}_{n_k}^{-1} - \left(\underline{X}_{id_k[p]}\underline{X}_{id_k[p]}^T\tau_p^2+\underline{R}_{n_k}\right)^{-1}\right]$$
$$\times\left(\underline{Y}_{[id_k]} - \underline{X}_{id_k[-p]}\underline{\beta}_{k[-p]}\right).$$
$$= \log(1-w) - \log(w)$$
$$+\ \frac{1}{2}\log\left(1+\tau_p^2\underline{X}_{id_k[p]}^T\underline{R}_{n_k}^{-1}\underline{X}_{id_k[p]}\right)$$
$$-\ \frac{1}{2}\left(\underline{Y}_{[id_k]} - \underline{X}_{id_k[-p]}\underline{\beta}_{k[-p]}\right)^T\underline{R}_{n_k}^{-1}\underline{X}_{id_k[p]}$$
$$\times\left[\frac{1}{\tau_p^2}+\underline{X}_{id_k[p]}^T\underline{R}_{n_k}^{-1}\underline{X}_{id_k[p]}\right]^{-1}\underline{X}_{id_k[p]}^T\underline{R}_{n_k}^{-1}$$
$$\times\left(\underline{Y}_{[id_k]} - \underline{X}_{id_k[-p]}\underline{\beta}_{k[-p]}\right),$$

where $\underline{R}_{n_k}$ is a $(M\times n_k)\times(M\times n_k)$ block diagonal matrix with $\underline{R}_k$ in each diagonal block.

(6) Full conditional distribution of w:

$$\pi\left(w|\ all\ others\right)\propto\pi\left(\underline{Z}|w\right)\pi\left(w\right)$$
$$\propto\left[\prod_{p=1}^{P}\prod_{k=1}^{K}w^{Z_{kp}}(1-w)^{1-Z_{kp}}\right]w^{a-1}(1-w)^{b-1}.$$

Therefore,

$$\pi\left(w|\ all\ others\right)$$
$$= Beta\left(a+\sum_{p=1}^{P}\sum_{k=1}^{K}Z_{kp},\ b+\sum_{p=1}^{P}\sum_{k=1}^{K}(1-Z_{kp})\right).$$

(7) Full conditional distribution of τ_p^2:

$$\pi\left(\tau_p^2|\ all\ others\right)$$
$$= \pi\left(\underline{\beta}|Z_{kp},\tau_p^2\right)\pi\left(\tau_p^2\right)$$
$$\propto\left[\prod_{k=1:K;Z_{kp}=1}\frac{1}{\sqrt{\tau_p^2}}exp\left\{-\frac{\beta_{kp}^2}{2\tau_p^2}\right\}\right]\left(\tau_p^2\right)^{-s_{p1}-1}exp\left\{-\frac{s_{p2}}{\tau_p^2}\right\}$$
$$\propto\left(\frac{1}{\tau_p^2}\right)^{\frac{\sum_{k=1}^{K}\mathbf{1}_{\{Z_{kp}=1\}}}{2}+s_{p1}+1}$$
$$\times\ exp\left\{-\frac{\sum_{k=1}^{K}\beta_{kp}^2\mathbf{1}_{\{Z_{kp}=1\}}+2s_{p2}}{2\tau_p^2}\right\}$$

$$\pi\left(\tau_p^2|\ all\ others\right)$$
$$= InvGamma\left(s_{p1}+\frac{\sum_{k=1}^{K}\mathbf{1}_{\{Z_{kp}=1\}}}{2},\ s_{p2}\right.$$
$$\left.+\ \frac{\sum_{k=1}^{K}\beta_{kp}^2\mathbf{1}_{\{Z_{kp}=1\}}}{2}\right).$$

(8) Full conditional distribution of $\underline{\Sigma}_k$:
We use the following one-to-one transformation mapping from $\{\underline{Y}_i, \underline{R}_k\}$ to $\{\underline{W}_i, \underline{\Sigma}_k\}$ ($i \in id_k$):

$$\underline{Y}_i = \underline{X}_i \underline{\beta}_{G_i} + \underline{D}_{G_i}^{-1} \underline{W}_i,$$

$$\underline{R}_k = \underline{D}_k^{-1} \underline{\Sigma}_k \underline{D}_k^{-1},$$

where $\sum_{i \in id_k} W_{im}^2 = 1$ for any $m = 1, \ldots, M$, and $\underline{D}_k$ is a diagonal matrix. Note, given $\underline{\beta}_{G_i}$, the step that draws $\underline{Y}_i$ implicitly draws $\underline{W}_i$ and $\underline{D}_k$ because $\underline{D}_{kj} = \left\{ \sum_{i \in id_k} \left(Y_{ij} - \underline{X}_{ij} \underline{\beta}_{G_i} \right)^2 \right\}^{-1}$, where $\underline{D}_{kj}$ is the jth element of $\underline{D}_k$ and $\underline{X}_{ij}$ is the jth row of $\underline{X}_i$.
Assuming the following candidate prior for $\underline{R}_k$:

$$\pi_2 \left(\underline{R}_k \right) \propto \left| \underline{R}_k \right|^{-\frac{M+1}{2}}, \ \underline{R}_k \in \mathcal{J},$$

the joint distribution of $\underline{Y}$ and $\underline{R}_k$ given $\underline{\beta}_k$ is:

$$\pi \left(\underline{Y}_{[id_k]}, \underline{R}_k | \underline{\beta}_k \right) \propto \left| \underline{R}_k \right|^{-\frac{n_k}{2}} \times \exp \left(-\frac{1}{2} \sum_{i \in id_k} \left(\underline{Y}_i - \underline{X}_i \underline{\beta}_{G_i} \right)^T \underline{R}_k^{-1} \left(\underline{Y}_i - \underline{X}_i \underline{\beta}_{G_i} \right) \right) \times \left| \underline{R}_k \right|^{-\frac{M+1}{2}}.$$

Since the transformation Jacobian of $\left(\underline{Y}_{[id_k]}, \underline{R}_k \right) \rightarrow \left(\underline{W}_{[id_k]}, \underline{\Sigma}_k \right)$ is $J = \left| \underline{D}_k \right|^{-(n_k+M+1)}$,

$$\pi \left(\underline{W}_{[id_k]}, \underline{\Sigma}_k | \beta_k \right) \propto \left| \underline{R}_k \right|^{-\frac{n_k+M+1}{2}} \times \exp \left(-\frac{1}{2} \sum_{i \in id_k} \underline{W}_i^T \underline{\Sigma}_k^{-1} \underline{W}_i \right) \times J$$

$\propto \left| \underline{\Sigma}_k \right|^{-\frac{n_k+M+1}{2}} \exp \left[-\frac{1}{2} tr \, \underline{\Sigma}_k^{-1} \underline{S}_k \right]$. Thus, the full conditional distribution of $\underline{\Sigma}_k$ follows an Inverse Wishart distribution:

$$\underline{\Sigma}_k | \underline{W}_{[id_k]}, \underline{\beta}_k \sim \text{InvWishart} \left(\upsilon_k, \underline{S}_k \right),$$

where $\upsilon_k = nk$, $\underline{S}_k = \left(\sum_{i \in id_k} \underline{W}_i \underline{W}_i^T \right)$, and $\underline{W}_{[id_k]} = \underline{D}_k \left(\underline{Y}_{[id_k]} - \underline{X}_{id_k} \underline{\beta}_{G_i} \right)$. Next, we translate $\underline{\Sigma}_k$ back to $\underline{R}_k$ and compute the acceptance probability α at iteration l+1 which is given by:

$$\alpha = \min \left\{ 1, \frac{\pi_2(\underline{R}_k^{(l)})}{\pi_2(\underline{R}_k)} \right\} = \min \left\{ 1, \left(\frac{\left| \underline{R}_k^{(l)} \right|}{\left| \underline{R}_k \right|} \right)^{-\frac{M+1}{2}} \right\} = \min \left\{ 1, \exp \left(\frac{M+1}{2} \left(log \left| \underline{R}_k \right| - log \left| \underline{R}_k^{(l)} \right| \right) \right) \right\}.$$

Finally, a summary of the steps in our MCMC sampling algorithm is given below:

(1) Initialization of $\underline{Y}_i$, $\underline{d}$, G_i, $\underline{\beta}_k$, $\underline{Z}_k$, w, τ_p^2, and $\underline{R}_k$.
(2) Generate $\underline{Y}_i$ from a Truncated Multivariate Normal distribution.
(3) Generate $\underline{d}$ from a Dirichlet distribution.
(4) Generate G_i from a Categorical distribution.
(5) Jointly generate $\underline{\beta}_k$ and $\underline{Z}_k$ from a Normal distribution and a Bernoulli distribution.
(6) Generate hyper-parameters of w, τ_p^2 with a beta distribution and inverse Gamma distribution, respectively.
(7) Generate a covariance matrix from a Wishart distribution and compute the corresponding correlation matrix. Accept the candidate $\underline{R}_k$ with an appropriate acceptance probability.
(8) Apply relabeling algorithm for parameters $\underline{Z}_k$, G_i, $\underline{d}$, $\underline{R}_k$ through the 1,…, K segments based on $\beta_{1p} < \beta_{2p} < \cdots < \beta_{Kp}$ ($p = 1, \ldots, P$).
(9) Save the posterior samples of $\underline{Y}_i$, $\underline{d}$, G_i, $\underline{\beta}_k$, $\underline{Z}_k$, w, τ_p^2, and $\underline{R}_k$ per MCMC iteration.

Appendix B. Monte Carlo simulation study

We present a simulation analysis employing a fractional factorial design to investigate the performance of our proposed model with respect to differing model, data, and specification factors across various conditions.

B.1. The fractional factorial design

Table B.1 presents seven experimentally manipulated factors and their levels in this Monte Carlo analysis: the number of subjects (X1: $N = 300$, $N = 650$, and $N = 1000$), the number of clusters (X2: $K = 2$, $K = 3$, and $K = 4$), the number of choices/stimuli per subject (X3: $m = 10$, $m = 15$, and $m = 20$), the number of independent variables (X4: $P = 8$, $P = 16$, and $P = 24$), the proportion of active independent variables (X5: $l = 1/4$, $l = 1/3$, and $l = 1/2$), the specification of the correlation matrix (X6: *Identity matrix (I), Same correlation matrix across segments (R),* and *Different multiple correlation matrices across segments (R_k)*), and the hyperparameters (a and b; see Eq. (6) in the paper) (X7: *Vague prior with $a = 5$ and $b = 5$*, *Intermediate prior with $a = 10$ and $b = 10$*, and *Informative prior with $a = 10$ and $b = 50$*). Note, to generate the correlation matrices for X6, we used *rcorrmatrix* command in the *clusterGeneration* R package. These factors with three levels were specified to reflect various conditions representing a variety of potential applications, as well as the flexibility of the proposed methodology in fitting a variety of different models and data specifications. Regarding the data generation procedure, for pre-specified N, K, m, P, l, R_k, and hyperparameter values of a and b, true $\underline{Z}_k$ and true $\underline{\beta}_k$ were generated from a Bernoulli distribution and normal distribution respectively. Finally, latent utilities $\underline{Y}_i$ were generated according to Eq. (2) in the paper, and the choices $\underline{C}_i$ were generated following if $Y_{im} > 0$, $C_{im} = 1$, otherwise $C_{im} = 0$.

We attempted to create a variety of empirical settings which would realistically test the performance of the proposed model. Naturally, we would expect that conditions with larger datasets, smaller number of segments, smaller numbers of parameters, and accurate informative priors would provide conditions where the model would be expected to show better performances. In contrast, alternative conditions with smaller datasets, larger number of segments, larger number of parameters and availability of vague priors only would likely be more challenging for this or any particular methodology.

Consistent with the past psychometric literature (e.g., DeSarbo, 1982; DeSarbo and Carroll, 1985; DeSarbo and Cron, 1988; Jedidi and DeSarbo, 1991) involving Monte Carlo testing of newly proposed methods, we created a collection of synthetic datasets as generated by use of an orthogonal fractional factorial design by manipulating independent factors reflecting different data, parameter, and model specification conditions. We employ a 3^7 fractional factorial design (SPSS ORTHOPLAN command; Adelman, 1962) to study the main effects of each factor.

Table B.1
Experimental design factors for the simulation.

Factor	Levels	Code
Number of subjects (X1)	$N = 300$ $N = 650$ $N = 1000$	1 2 3
Number of segments (X2)	$K = 2$ $K = 3$ $K = 4$	1 2 3
Number of choices/brands per subject (X3)	$m = 10$ $m = 15$ $m = 20$	1 2 3
Number of independent variables (X4)	$P = 8$ $P = 16$ $P = 24$	1 2 3
Proportion of active independent variables (X5)	$l = 1/4$ $l = 1/3$ $l = 1/2$	1 2 3
Correlation matrices specification across segments (X6)[a]	$R_k = I$ $R_k = R$ $R_k = R_k$	1 2 3
Hyperparameters (a and b) of Beta distribution of w (X7)[b]	*Vague prior* *Intermediate* *Informative prior*	1 2 3

[a] I refers Identity matrix; R refers one identical correlation matrix across segments; R_k refers K different correlation matrices across segments.
[b] "*Vague*" sets $a = 5$ and $b = 5$ (mean = 0.5, standard deviation = 0.15); "*Intermediate*" $a = 10$ and $b = 10$ (mean = 0.5, standard deviation $v = 0.1$); "*Informative*" $a = 10$ and $b = 50$ (mean = 0.167, standard deviation = 0.05).

Table B.2
Average values of dependent measures across three levels of seven factors.

		X1	X2	X3	X4	X5	X6	X7
Recovery of data/choices	1	.969	.981	.959	.968	.968	.973	.973
	2	.948	.955	.974	.965	.973	.968	.960
	3	.972	.954	.956	.956	.948	.948	.957
	Total	.963	.963	.963	.963	.963	.963	.963
% of correct independent variable selected	1	.937	.961	.944	.952	.958	.964	.943
	2	.915	.958	.934	.949	.958	.950	.928
	3	.971	.903	.944	.921	.907	.909	.951
	Total	.941	.941	.941	.941	.941	.941	.941
% of wrong independent variable selected	1	.950	.958	.913	.936	.933	.973	.942
	2	.923	.940	.952	.929	.939	.928	.929
	3	.962	.937	.971	.971	.963	.934	.964
	Total	.945	.945	.945	.945	.945	.945	.945
Membership hit rate	1	.948	.961	.930	.930	.940	.946	.946
	2	.941	.954	.948	.948	.936	.950	.936
	3	.940	.915	.952	.952	.954	.934	.948
	Total	.943	.943	.943	.943	.943	.943	.943
RMSE coefficient recovery	1	1.058	1.229	.906	1.194	1.203	1.366	1.212
	2	1.090	.892	1.088	.960	.998	.881	.711
	3	.917	.943	1.071	.911	.864	.818	1.141
	Total	1.022	1.022	1.022	1.022	1.022	1.022	1.022
CPU time (minutes)	1	240	728	490	470	506	873	809
	2	472	515	505	721	637	515	473
	3	1184	652	901	704	752	507	614
	Total	632	632	632	632	632	632	632

The six dependent measures computed were:

1. Recovery of choices in data: the proportion of correct classifications of choices for observations in the data.
2. The percentage of correct independent variable selected:
$$\frac{\textit{the number of correctly selected active variables}}{\textit{the number of all active variables}} \times 100$$
3. The percentage of wrong independent variable selected:
$$\frac{\textit{the number of wrongly selected variables}}{\textit{the number of all inactive variables}} \times 100$$
4. The root mean squared error (RMSE) between the actual and recovered segment level coefficients $\underline{\beta}_k$;
5. The segment membership hit rate:
$$\frac{\textit{the number of correct memberships predicted}}{\textit{total sample size}}$$
6. CPU time for 20,000 iterations (i.e., 10,000 burn-in and 10,000 sampling).

In this Monte Carlo analysis, we used a total of 20,000 iterations, and after discarding the first 10,000 iterations as burn-in, the next 10,000 iterations were collected for parameter estimation and choice prediction per trial. To track convergence, we employed

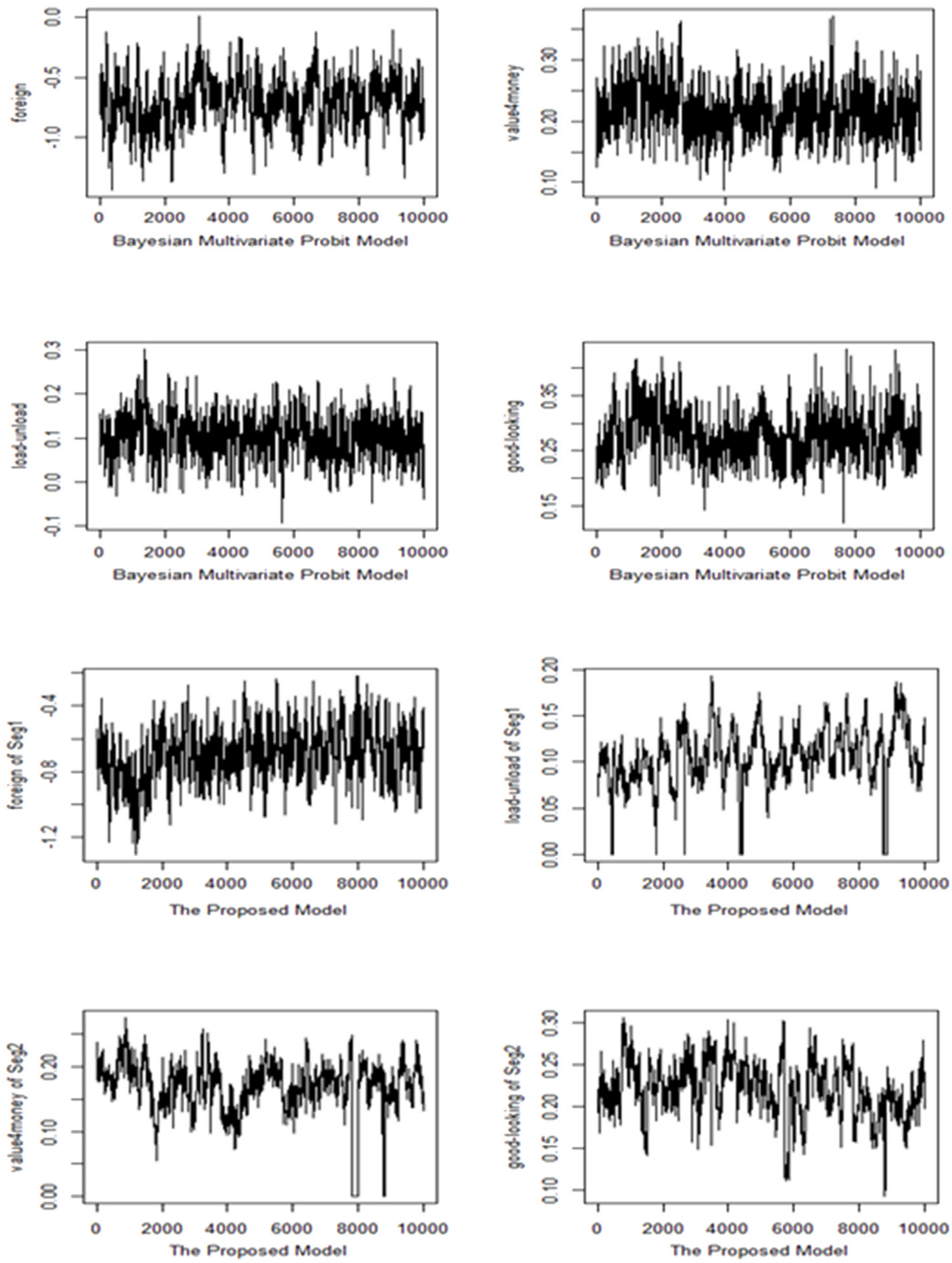


Fig. C.1. Sample trace plots of coefficients of Bayesian multivariate probit model (*bayesm*, Rossi (2017)) and the proposed model in the application study.

trace plots which exhibited consistent patterns after the burn-in period, and the Geweke (1992) diagnostic.

B.2. Monte Carlo study results

Table B.2 presents the average performances for six dependent measures across three levels of seven factors. In specific, the proposed model provides good performance for the recovery of data/choices (96.3%), proportion of correctly selected independent variables (94.1%), proportion of incorrectly selected independent variables (94.5%), and membership hit rate (94.3%). This shows that the proposed model successfully recovers the true parameters and choices/data (robustness of the performance will be discussed shortly below). The RMSE of true coefficients recovery and CPU time are 1.022 and 632 min for 20,000 iterations.

Consistent with (DeSarbo, 1982; DeSarbo & Carroll, 1985; DeSarbo & Cron, 1988; Jedidi & DeSarbo, 1991), we performed linear regression analysis on each of the six dependent variables after coding the seven independent factors as dummy variables to explore main effects (akin to ANOVA). Note, we used a middle level among three levels of each factor as reference level. A primary indication of methodological robustness would be to find *non-significance* of each regression model (F-test) as well as independent factors (t-tests) suggesting that the proposed MCMC estimation procedure is not significantly affected by data, prior, or model specifications. The Table B.3 reports the result statistics summary by multiple linear regression analyses with respect to the six dependent measures. Regarding the linear regression results using recovery of choices (dependent measure 1) as the dependent variable and the seven factors as independent variables, the overall linear model is not significant (p-value is 0.248, adj. $R^2 = 0.208$). Indeed, the corresponding table of coefficients indicates that none of the factors' levels are significant at the 0.05 level. The same holds true in the three analysis results using dependent measures 2, 3 and 4 (i.e., finding correct independent variables, finding wrong

Table B.3
Linear model result table.

	Recovery of choices	% of correct IVs	% of wrong IVs	RMSE of beta recovery	Segment membership hit rate	CPU time
Model: F-test (p-value)	1.49 (0.25)	1.34 (0.31)	0.94 (0.55)	0.76 (0.69)	1.52 (0.24)	2.5 (0.06)
Adj. R^2	0.208	0.153	−0.034	−0.146	0.219	0.447
X1_D1: t-stat (p-value)	1.6 (0.14)	0.74 (0.47)	0.87 (0.40)	−0.1 (0.92)	0.48 (0.64)	−1.07 (0.31)
X1_D2: t-stat (p-value)	1.82 (0.09)	1.91 (0.08)	1.27 (0.23)	−0.55 (0.60)	−0.7 (0.95)	3.27 (0.01)*
X2_D1: t-stat (p-value)	1.9 (0.08)	0.09 (0.93)	0.57 (0.58)	1.1 (0.31)	0.46 (0.65)	0.98 (0.35)
X2_D2: t-stat (p-value)	−0.07 (0.95)	−1.88 (0.08)	−0.11 (0.92)	0.16 (0.87)	−2.72 (0.02)*	0.63 (0.54)
X3_D1: t-stat (p-value)	−1.08 (0.30)	0.33 (0.75)	−1.27 (0.23)	−0.57 (0.58)	−1.22 (0.24)	−0.07 (0.95)
X3_D2: t-stat (p-value)	−1.34 (0.21)	0.35 (0.74)	0.64 (0.54)	−0.05 (0.96)	0.31 (0.76)	1.82 (0.09)
X4_D1: t-stat (p-value)	0.2 (0.85)	0.1 (0.93)	0.22 (0.83)	0.74 (0.47)	−1.25 (0.24)	−1.15 (0.27)
X4_D2: t-stat (p-value)	−0.64 (0.54)	−0.97 (0.35)	1.38 (0.19)	−0.16 (0.88)	0.33 (0.75)	−0.08 (0.94)
X5_D1: t-stat (p-value)	−0.42 (0.69)	0 (1.00)	−0.22 (0.83)	0.65 (0.53)	0.26 (0.80)	−0.6 (0.56)
X5_D2: t-stat (p-value)	−1.86 (0.09)	−1.73 (0.11)	0.78 (0.45)	−0.42 (0.68)	1.2 (0.25)	0.53 (0.61)
X6_D1: t-stat (p-value)	0.42 (0.69)	0.475 (0.64)	1.47 (0.17)	1.54 (0.15)	−0.33 (0.75)	1.65 (0.13)
X6_D2: t-stat (p-value)	−1.44 (0.17)	−1.41 (0.18)	0.18 (0.86)	−0.2 (0.85)	−1.14 (0.28)	−0.04 (0.97)
X7_D1: t-stat (p-value)	0.95 (0.36)	0.52 (0.61)	0.42 (0.68)	1.59 (0.14)	0.67 (0.51)	1.54 (0.15)
X7_D2: t-stat (p-value)	−.212 (0.84)	0.78 (0.45)	1.12 (0.28)	1.36 (0.20)	0.78 (0.45)	0.65 (0.53)

Note: For all dummy variables coding, we choose the middle level (e.g., $N = 650$, $K = 3$, $m = 15$, etc.) as a reference group;
* Refers significant at 0.05 level.

independent variables, and RMSE for recovery of true β_k). The multiple ANOVA tests show that the overall linear models are not significant (p-value $= 0.311$, adj. $R^2 = 0.153$ for dependent measure 2; p-value $= 0.55$, adj. $R^2 = -0.034$ for dependent measure 3; p-value $= 0.69$, adj. $R^2 = -0.146$ for dependent measure 4). In sum, none of the factors or none of their coefficients (levels of the factors) are significant at the 0.05 level for dependent measures 1 to 4.

For dependent measure 5 (segment membership hit rate), the ANOVA test indicates that the overall model is also not significant (p-value $= 0.236$, adj. $R^2 = 0.219$), but the X2_D2 factor in the Table B.3 (i.e., a dummy factor between $K = 3$ and $K = 4$ of X2) is significant at the 0.05 level. Note, the hit rate for $K = 4$ is still good with 91% (note, 96% for $K = 2$; 95% for $K = 3$). Last, for dependent measure 6 (CPU time), the overall model is marginally significant (p-value $= 0.06$, adj $R^2 = 0.447$) but only X1_D2 factor in the Table B.3 (i.e., a dummy factor between $N = 650$ and $N = 1000$) is significant at 0.05 level. This result also makes sense, as increasing number of samples should need more computing time. Overall, this Monte Carlo simulation study demonstrates the proposed new Bayesian methodology successfully recovers various true parameters, and appears robust with respect to these seven factors and their levels.

Appendix C. Sample trace plots for checking convergence

see Fig. C.1.

References

Adelman, M. A. (1962). *The economics of petroleum supply: Papers by MA Adelman*. (pp. 1962–1993). MIT press.

Alba, J. W., & Hutchinson, J. W. (1987). Dimensions of consumer expertise. *Journal of Consumer Research*, *13*(4), 411–454.

Andrews, J. L., & McNicholas, P. D. (2014). Variable selection for clustering and classification. *Journal of Classification*, *31*(2), 136–153.

Arminger, G., Clogg, C. C., & Cheng, T. (2000). 1. Regression analysis of multivariate binary response variables using rasch-type models and finite-mixture methods. *Sociological Methodology*, *30*(1), 1–26.

Benati, S., & García, S. (2014). A mixed integer linear model for clustering with variable selection. *Computers & Operations Research*, *43*, 280–285.

Bhat, C. R. (2017). Incorporating observed and unobserved heterogeneity in urban work travel mode choice modeling. Working Paper, University of Texas, Austin TX.

Bontemps, D., & Toussile, W. (2013). Clustering and variable selection for categorical multivariate data. *Electronic Journal of Statistics*, *7*, 2344–2371.

Bouveyron, C., & Brunet-Saumard, C. (2014). Discriminative variable selection for clustering with the sparse Fisher-EM algorithm. *Computational Statistics*, *29*(3–4), 489–513.

Celeux, G., Forbes, F., Robert, C. P., & Titterington, D. M. (2006). Deviance information criteria for missing data models. *Bayesian Analysis*, *1*(4), 651–673.

Chib, S., & Greenberg, E. (1998). Analysis of multivariate probit models. *Biometrika*, *85*(2), 347–361.

Cowley, E., & Mitchell, A. A. (2003). The moderating effect of product knowledge on the learning and organization of product information. *Journal of Consumer Research*, *30*(3), 443–454.

Cozzini, A., Jasra, A., Montana, G., & Persing, A. (2014). A Bayesian mixture of lasso regressions with t-errors. *Computational Statistics & Data Analysis*, *77*, 84–97.

Dean, N., & Raftery, A. E. (2010). Latent class analysis variable selection. *Annals of the Institute of Statistical Mathematics*, *62*(1), 11–35.

Dellaert, B. G., & Stremersch, S. (2005). Marketing mass-customized products: Striking a balance between utility and complexity. *Journal of Marketing Research*, *42*(2), 219–227.

DeSarbo, W. S. (1982). GENNCLUS: New models for general nonhierarchical clustering analysis. *Psychometrika*, *47*(4), 449–475.

DeSarbo, W. S., & Carroll, J. D. (1985). Three-way metric unfolding via alternating weighted least squares. *Psychometrika*, *50*(3), 275–300.

DeSarbo, W. S., & Cron, W. L. (1988). A maximum likelihood methodology for clusterwise linear regression. *Journal of Classification*, *5*(2), 249–282.

DeSarbo, W. S., Fong, D. K., Liechty, J., & Coupland, J. C. (2005). Evolutionary preference/utility functions: A dynamic perspective. *Psychometrika*, *70*(1), 179–202.

DeSarbo, W. S., Park, J., & Rao, V. R. (2011). Deriving joint space positioning maps from consumer preference ratings. *Marketing Letters*, *22*(1), 1–14.

DeSoete, G., & DeSarbo, W. S. (1991). A latent class probit model for analyzing pick any/n data. *Journal of Classification*, *8*, 45–63.

Dubé, J. P. (2004). Multiple discreteness and product differentiation: Demand for carbonated soft drinks. *Marketing Science*, *23*(1), 66–81.

Fan, J., & Li, R. (2001). Variable selection via nonconcave penalized likelihood and its oracle properties. *Journal of the American Statistical Association*, *96*(456), 1348–1360.

Feldman, J. M., & Lynch, J. G. (1988). Self-generated validity and other effects of measurement on belief, attitude, intention, and behavior. *Journal of Applied Psychology*, *73*(3), 421.

Freese, J., & Long, J. S. (2006). *Regression models for categorical dependent variables using stata*. College Station: Stata Press.

Gelman, A., Carlin, J. B., Stern, H. S., Dunson, D. B., Vehtari, A., & Rubin, D. B. (2014). *Bayesian data analysis*. (3rd ed.). Chapman and Hall/CRC.

Gelman, A., & Rubin, D. B. (1992). Inference from iterative simulation using multiple sequences. *Statistical Science*, *7*(4), 457–511.

George, E. I., & McCulloch, R. E. (1993). Variable selection via Gibbs sampling. *Journal of the American Statistical Association*, *88*(423), 881–889.

George, E., & McCulloch, R. (1997). Approaches for Bayesian variable selection. *Statistica Sinica*, *7*(2), 339–373.

Geweke, J. (1992). Evaluating the accuracy of sampling-based approaches to calculating posterior moments. In J. M. Bernado, J. O. Berger, A. P. Dawid, & A. F. M Smith (Eds.), *Bayesian statistics, Vol. 4*. Oxford, UK: Clarendon Press.

Ghosh, J., Herring, A. H., & Siega-Riz, A. M. (2011). Bayesian variable selection for latent class models. *Bioemtrics*, *67*(3), 917–925.

Gilbride, T. J., Allenby, G. M., & Brazell, J. D. (2006). Models for heterogeneous variable selection. *Journal of Marketing Research*, *43*(3), 420–430.

Green, P. E., & Srinivasan, V. (1978). Conjoint analysis in consumer research: issues and outlook. *Journal of Consumer Research*, *5*(2), 103–123.

Hauser, J. R. (2014). Consideration-set heuristics. *Journal of Business Research*, *67*, 1688–1699.

Houston, D. A., & Sherman, S. J. (1995). Cancellation and focus: The role of shared and unique features in the choice process. *Journal of Experimental Social Psychology*, *31*(4), 357–378.
Hutchinson, J. W., & Alba, J. W. (1991). Ignoring irrelevant information: Situational determinants of consumer learning. *Journal of Consumer Research*, *18*(3), 325–345.
Isen, A. M. (1993). Positive affect and decision making. In M. Lewis, & J. Haviland (Eds.), *Handbook of emotions* (pp. 261–273). New York: Guilford.
Isen, A. M., Daubman, K. A., & Nowicki, G. P. (1987). Positive affect facilitates creative problem solving. *Journal of Personality and Social Psychology*, *52*(6), 1122–1131.
Ishwaran, H., & Rao, J. S. (2011). Consistency of spike and slab regression. *Statistics & Probability Letters*, *81*(12), 1920–1928.
Jedidi, K., & DeSarbo, W. S. (1991). A stochastic multidimensional scaling procedure for the spatial representation of three-mode, three-way pick any/J data. *Psychometrika*, *56*(3), 471–494.
Jeffreys, Harold (1961). *Theory of probability*. (3rd ed.). Oxford: Clarendon Press.
Johnson, R. M. (1987). Markov chain monte carlo maximum likelihood. In *Sawtooth software conference on perceptual mapping, conjoint analysis, and computer interviewing*.
Johnson, K. E., & Mervis, C. B. (1997). Effects of varying levels of expertise on the basic level of categorization. *Journal of Experimental Psychology: General*, *126*(3), 248.
Kamakura, W. A., Kim, B. D., & Lee, J. (1996). Modeling preference and structural heterogeneity in consumer choice. *Marketing Science*, *15*(2), 152–172.
Khalili, A., & Chen, J. (2007). Variable selection in finite mixture of regression models. *Journal of the American Statistical Association*, *102*(479), 1025–1038.
Kim, J., Allenby, G. M., & Rossi, P. E. (2002). Modeling consumer demand for variety. *Marketing Science*, *21*(3), 229–250.
Kim, S., Fong, D. K., & DeSarbo, W. S. (2012). Model-based segmentation featuring simultaneous segment-level variable selection. *Journal of Marketing Research*, *49*(5), 725–736.
Kim, S., Tadesse, M. G., & Vannucci, M. (2006). Variable selection in clustering via Dirichlet process mixture models. *Biometrika*, 877–893.
Kunda, Z. (1990). The case for motivated reasoning. *Psychological Bulletin*, *108*(3), 480.
Leong, W., & Hensher, D. A. (2012). Embedding decision heuristics in discrete choice model: A review. *Transport Reviews*, *32*(3), 313–331.
Li, Y., & Ansari, A. (2014). A Bayesian semiparametric approach for endogeneity and heterogeneity in choice models. *Management Science*, *60*(5), 1161–1179.
Li, S., Sun, B., & Wilcox, R. T. (2005). Cross-selling sequentially ordered products: An application to consumer banking services. *Journal of Marketing Research*, *42*(2), 233–239.
Liechty, J., Liechty, M., & Müller, P. (2004). Bayesian correlation estimation. *Biometrika*, *91*(1), 1–14.
Liu, X., & Daniels, M. J. (2006). A new algorithm for simulating a correlation matrix based on parameter expansion and reparameterization. *Journal of Computational and Graphical Statistics*, *15*(4), 897–914.
Liu, J. S., & Wu, Y. N. (1999). Parameter expansion for data augmentation. *Journal of the American Statistical Association*, *94*(448), 1264–1274.
Marin, J. M., & Robert, C. (2007). *Bayesian core: A practical approach to computational Bayesian statistics*. New York: Springer Science & Business Media.
Miller, K. E., & Ginter, J. L. (1979). An investigation of situational variation in brand choice behavior and attitude. *Journal of Marketing Research*, *16*(1), 111–123.
Mitchell, T. J., & Beauchamp, J. J. (1988). Bayesian variable selection in linear regression. *Journal of the American Statistical Association*, *83*(404), 1023–1032.
Okada, K., & Lee, M. D. (2016). A bayesian approach to modeling group and individual differences in multidimensional scaling. *Journal of Mathematical Psychology*, *70*(1), 35–44.
Park, J., DeSarbo, W. S., & Liechty, J. (2008). A hierarchical Bayesian multidimensional scaling methodology for accommodating both structural and preference heterogeneity. *Psychometrika*, *73*(4), 451–472.
Park, J., Rajagopel, P., Dillon, W., & Chaiy, S. (2017). A multidimensional unfolding model for assessing dimensionality selection in product positioning applications. Working Paper. Cox School of Business, Southern Methodist University, Dallas TX.
Payne, J. W., Bettman, J. R., & Johnson, E. J. (1993). *The adaptive decision maker*. Cambridge: Cambridge University Press.
Payne, J. W., Bettman, J. R., & Luce, M. F. (1996). When time is money: Decision behavior under opportunity-cost time pressure. *Organizational Behavior and Human Decision Processes*, *66*(2), 131–152.
Petty, R., & Cacioppo, J. (2012). *Communication and persuasion: Central and peripheral routes to attitude change*. New York: Springer Science & Business Media.
Petty, R. E., Cacioppo, J. T., & Goldman, R. (1981). Personal involvement as a determinant of argument-based persuasion. *Journal of Personality and Social Psychology*, *41*(5), 847.
Raftery, A. E., & Dean, N. (2006). Variable selection for model-based clustering. *Journal of the American Statistical Association*, *101*(473), 168–178.
Ritter, G. (2014). *Robust cluster analysis and variable selection*. New York: CRC Press.
Rossi, P. E. (2017). Bayesian Inference for Marketing/Micro-econometrics. *bayesm* in R Package.
Rossi, P. E., Allenby, G. M., & McCulloch, R. E. (2005). *Bayesian statistics and marketing*. (pp. 1–368). New York: Wiley.
Rousseeuw, P. J., & Molenberghs, G. (1994). The shape of correlation matrices. *The American Statistician*, *48*(4), 276–279.
Sperrin, M., Jaki, T., & Wit, E. (2010). Probabilistic relabelling strategies for the label switching problem in Bayesian mixture models. *Statistics and Computing*, *20*(3), 357–366.
Spiegelhalter, D. J., Best, N. G., Carlin, B. P., & Van Der Linde, A. (2002). Bayesian measures of model complexity and fit. *Journal of the Royal Statistical Society. Series B. Statistical Methodology*, *64*(4), 583–639.
Steele, R. J., & Raftery, A. E. (2010). Performance of Bayesian model selection criteria for gaussian mixture models. In M.-H. Chen (Ed.), *Frontiers of statistical decision making and bayesian analysis* (pp. 113–130). New York: Springer.
Sujan, M. (1985). Consumer knowledge: Effects on evaluation strategies mediating consumer judgments. *Journal of Consumer Research*, *12*(1), 31–46.
Tadesse, M. G., Sha, N., & Vannucci, M. (2005). Bayesian variable selection in clustering high-dimensional data. *Journal of the American Statistical Association*, *100*(470), 602–617.
Tibshirani, R. (1996). Regression shrinkage and selection via the lasso. *Journal of the Royal Statistical Society. Series B. Statistical Methodology*, *58*(1), 267–288.
Van Dyk, D. A., & Park, T. (2008). Partially collapsed Gibbs samplers: Theory and methods. *Journal of the American Statistical Association*, *103*(482), 790–796.
Vicente, K. J., & Wang, J. H. (1998). An ecological theory of expertise effects in memory recall. *Psychological Review*, *105*(1), 33.
Wedel, M., Kamakura, W., et al. (2000). *Market segmentation: Conceptual and methodological foundations*. (2nd ed.). New York: Springer-Verlag.
Wedel, M., Kamakura, W., Arora, N., Bemmaor, A., Chiang, J., Elrod, T., et al. (1999). Discrete and continuous representations of unobserved heterogeneity in choice modeling. *Marketing Letters*, *10*(3), 219–232.
White, A., Wyse, J., & Murphy, T. B. (2016). Bayesian variable selection for latent class analysis using a collapsed Gibbs sampler. *Statistics and Computing*, *26*(1–2), 511–527.
Wright, P., & Weitz, B. (1977). Time horizon effects on product evaluation strategies. *Journal of Marketing Research*, *14*(4), 429–443.
Yang, C. (2005). Modeling multiple sources of heterogeneity in mode choice models. *Journal of Eastern Asia Society for Transportation Studies*, *6*, 2031–2045.