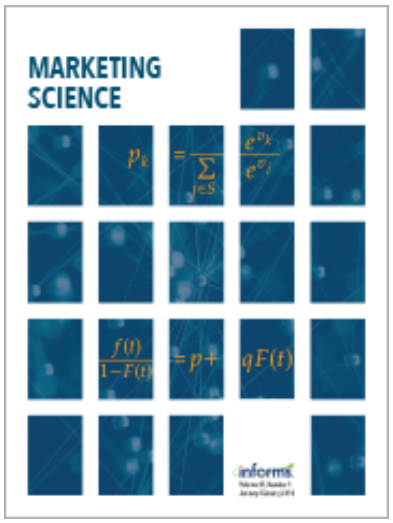
This article was downloaded by: [154.59.124.38] On: 30 May 2021, At: 08:47
Publisher: Institute for Operations Research and the Management Sciences (INFORMS)
INFORMS is located in Maryland, USA



Marketing Science

Publication details, including instructions for authors and subscription information:
http://pubsonline.informs.org

Explaining Preference Heterogeneity with Mixed Membership Modeling

Marc R. Dotson, Joachim Büschken, Greg M. Allenby

To cite this article:
Marc R. Dotson, Joachim Büschken, Greg M. Allenby (2020) Explaining Preference Heterogeneity with Mixed Membership Modeling. Marketing Science 39(2):407-426. https://doi.org/10.1287/mksc.2019.1185

Full terms and conditions of use: https://pubsonline.informs.org/Publications/Librarians-Portal/PubsOnLine-Terms-and-Conditions

This article may be used only for the purposes of research, teaching, and/or private study. Commercial use or systematic downloading (by robots or other automatic processes) is prohibited without explicit Publisher approval, unless otherwise noted. For more information, contact permissions@informs.org.

The Publisher does not warrant or guarantee the article's accuracy, completeness, merchantability, fitness for a particular purpose, or non-infringement. Descriptions of, or references to, products or publications, or inclusion of an advertisement in this article, neither constitutes nor implies a guarantee, endorsement, or support of claims made of that product, publication, or service.

Copyright © 2020, INFORMS

Please scroll down for article—it is on subsequent pages

With 12,500 members from nearly 90 countries, INFORMS is the largest international association of operations research (O.R.) and analytics professionals and students. INFORMS provides unique networking and learning opportunities for individual professionals, and organizations of all types and sizes, to better understand and use O.R. and analytics tools and methods to transform strategic visions and achieve better outcomes.
For more information on INFORMS, its publications, membership, or meetings visit http://www.informs.org

http://pubsonline.informs.org/journal/mksc

MARKETING SCIENCE
Vol. 39, No. 2, March–April 2020, pp. 407–426
ISSN 0732-2399 (print), ISSN 1526-548X (online)

Explaining Preference Heterogeneity with Mixed Membership Modeling

Marc R. Dotson,[a] Joachim Büschken,[b] Greg M. Allenby[c]

[a] Marriott School of Business, Brigham Young University, Provo, Utah 84602; [b] Catholic University of Eichstätt-Ingolstadt, 85072 Eichstätt, Germany; [c] Fisher College of Business, Ohio State University, Columbus, Ohio 43210

Contact: marc.dotson@byu.edu, http://orcid.org/0000-0001-7846-3660 (MRD); joachim.bueschken@kuei.de, http://orcid.org/0000-0002-9673-8928 (JB); allenby.1@osu.edu, http://orcid.org/0000-0001-9759-0067 (GMA)

Received: February 16, 2017
Revised: January 18, 2018; March 5, 2019; May 29, 2019
Accepted: June 7, 2019
Published Online in Articles in Advance: March 3, 2020

https://doi.org/10.1287/mksc.2019.1185

Copyright: © 2020 INFORMS

Abstract. Understanding the drivers of preferences requires models and covariates that explain their cross-sectional variation in models of demand. In this paper, we develop an integrated model of choice and covariates where parameters of the covariate model serve as regressors to explain preference heterogeneity in the choice model. We investigate alternative models to deal with many potential covariates and find that a grade of membership model provides empirical and conceptual advantages in explaining preferences. Our proposed model is illustrated with two conjoint data sets.

History: Yuxin Chen served as the senior editor and Carl Mela served as associate editor for this article.
Supplemental Material: Data and the online appendix are available at https://doi.org/10.1287/mksc.2019.1185.

Keywords: choice models • mixed membership models • hierarchical Bayes • grade of membership • preference heterogeneity

1. Introduction

Marketing researchers have increasing access to data that can potentially explain preference heterogeneity. Such data are often discrete, including binary responses to select-all-that-apply questions, fixed-point rating scales, records of consumer behavior, and text, and are useful if they can point to individual-level differences in preference. However, these data often require models for their interpretation because the raw data can be prone to confounding effects (Rossi et al. 2001) or may require summarization to provide greater meaning and context (Büschken and Allenby 2016). Hierarchical Bayesian (HB) choice models that incorporate preference heterogeneity are effective for determining what products consumers prefer. However, these models have proven less successful at explaining and interpreting preference heterogeneity. This paper explores a new model to incorporate discrete covariates into a choice model with heterogeneous preferences.

Including covariates to explain preferences in choice models can be problematic because of the resulting dimensionality of the coefficient matrix relating covariates to model parameters. For example, relating 10 part-worths in a conjoint study to 20 covariates in a model of heterogeneity requires at least a 10×20 coefficient matrix for the simplest data, which can be too demanding for even a moderately sized survey with 500 respondents. We investigate integrating a choice model with a covariate model to reduce the dimensionality and improve the interpretation of the coefficient matrix relating covariates and part-worths. The reduction in dimensionality comes from using parameters from the covariate model as regressors in the choice model rather than using the covariates themselves.

We develop a new HB choice model that relates individual-level part-worths to unobserved, individual-level membership vectors from a grade of membership (GoM) model to discrete covariates (Woodbury et al. 1978, Erosheva et al. 2007). The GoM model is related to the latent Dirichlet allocation (LDA) model, which serves as a touchstone within text analysis and topic modeling (Blei et al. 2003). Both LDA and GoM are part of a larger class of models known as mixed membership models that provide individual-level, low-dimensional representations of discrete data (Airoldi et al. 2014). This is accomplished by identifying patterns within the discrete data across respondents along with how each respondent relates to these patterns. We apply our model to two conjoint data sets and find that the integrated choice and GoM model improves predictive fit relative to either directly using the covariates as regressors or using competing models of heterogeneity.

This paper also contributes to efforts at using mixed membership models in marketing. The application of this class of models to marketing contexts is still in development. However, the dimension reduction inherent in these models is appealing. Extant research has focused on LDA models, using product reviews and online forums to inform market structure

(Lee and Bradlow 2011, Netzer et al. 2012) and to identify preferences for product features (Archak et al. 2011). More recently, Tirunillai and Tellis (2014) used LDA to conduct brand analysis, Jacobs et al. (2016) employed LDA to help predict purchases, Büschken and Allenby (2016) developed a sentence-constrained supervised LDA to better predict review ratings, Toubia and Netzer (2017) used topic models to explore idea generation, and Ansari et al. (2018) developed a scalable recommendation system based on supervised topic modeling. However, mixed membership models have yet to be employed in the context of choice modeling. We believe this paper provides an important step in this regard.

The remainder of this paper will be organized as follows. We specify our model in Section 2. In Section 3, we explore the boundary conditions of the proposed model relative to competing models by way of an extensive simulation study. We detail our two empirical applications in Section 4. In Section 5, we compare results from our proposed model and alternative models. We discuss implications of and extensions to this research in Section 6.

2. Model Specification

We detail the components of our proposed model before specifying their integration. First, we briefly review HB choice models and the use of covariates in explaining preference heterogeneity. Second, we outline the class of mixed membership models generally and the GoM model specifically, along with a comparison with related and competing covariate models. Finally, we specify the integrated choice and GoM model.

2.1. Hierarchical Bayesian Choice Model

Hierarchical Bayesian choice models enable both individual and aggregate-level preference parameter estimation, even in the presence of few choices per individual, by sharing information across individuals through the model of heterogeneity (Rossi and Allenby 2003, Rossi et al. 2005). Market simulators make use of individual-level preference estimates to forecast the results of potential product policies, whereas aggregate-level preference estimates are used to explain individual-level preference heterogeneity and inform targeting and promotion decisions.

The likelihood in an HB choice model is typically assumed to be a multinomial logit model such that the probability of individual n for choice task h choosing product alternative p is a function of the M attribute levels $x_{p,h}$ that compose alternative p for choice task h and individual n's M attribute-level part-worths β_n:

$$\Pr\left(y_{n,h} = p|\beta_n\right) = \frac{\exp\left(x'_{p,h}\beta_n\right)}{\sum_{p=1}^{P}\exp\left(x'_{p,h}\beta_n\right)}, \tag{1}$$

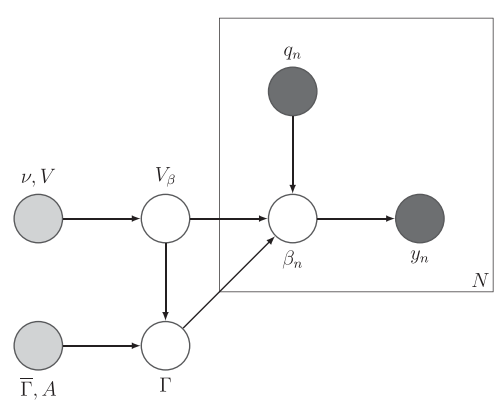
where there are P alternatives. The distribution of heterogeneity, or upper level, models preference heterogeneity across the individual-level β_n's. This model of heterogeneity is typically assumed to be multivariate normal (MVN) and, in the presence of covariates to part-worths for individual n, can be written down as

$$\beta_n = \Gamma q_n + \xi_n, \quad \xi_n \sim \mathrm{N}\left(0, V_\beta\right), \tag{2}$$

where q_n is a vector of observed covariates of length J, and Γ is the $J \times M$ coefficient matrix relating variation across $N \times J$ covariates Q to variation across $N \times M$ part-worths B. A simple variant of this model is given by $q_n = 1$ and $J = 1$ so that Γ contains an intercept only, resulting in $\beta_n \sim \mathrm{MVN}(\bar{\beta}, V_\beta)$. In the model with covariates, the prior expectation to β_n is Γq_n and, thus, heterogeneous. Information is shared across respondents through Γ and the $M \times M$ heterogeneity covariance matrix V_β (Rossi et al. 2005). In the intercept-only model, the prior expectation to β_n is homogeneous across respondents ($\bar{\beta}$), resulting in shrinkage of individual-level estimates to a common mean. Thus, in order to better explain preference heterogeneity, covariates are incorporated into the model to potentially allow for differences in prior expectations to the β_n and, in this sense, distinct points of shrinkage of their a posteriori estimates.

The directed acyclic graph (DAG) in Figure 1 provides a visual representation of the HB choice model with observed covariates. Plates in this DAG represent replications of variables, white nodes represent parameters, grey nodes represent fixed hyperparameters, and black nodes represent data. We use ν and V as degrees of freedom and the $M \times M$ scale matrix for a conjugate inverse Wishart prior on the heterogeneity covariance matrix V_β, respectively; $\bar{\Gamma}$ is the $J \times M$ mean, and A is the $J \times J$ precision matrix

Figure 1. Hierarchical Bayesian Choice Model



Notes. Plates in this DAG represent replications of variables, white nodes represent parameters, grey nodes represent fixed hyperparameters, and black nodes represent data. The fixed design matrix X is not shown to reduce clutter.

for a conjugate normal prior on the coefficient matrix Γ; β_n is a vector of M part-worth coefficients, y_n is a vector of H choices, and q_n is a vector of J covariates. Following Figure 1, the joint posterior distribution of the standard HB choice model is as follows, with the fixed design matrix X suppressed to reduce clutter:

$$p(B,\Gamma,V_\beta|Y,Q,\overline{\Gamma},A,\nu,V) \propto \left[\prod_{n=1}^{N} p(y_n|\beta_n)p(\beta_n|q_n,\Gamma,V_\beta)\right]p(\Gamma|V_\beta,\overline{\Gamma},A)p(V_\beta|\nu,V), \tag{3}$$

where Y is an $N \times H$ matrix of choices, $\prod_{n=1}^{N} p(y_n|\beta_n)$ is the likelihood, $\prod_{n=1}^{N} p(\beta_n|q_n,\Gamma,V_\beta)$ is the distribution of heterogeneity, and $p(\Gamma|V_\beta,\overline{\Gamma},A)$ and $p(V_\beta|\nu,V)$ are priors (Rossi et al. 2005).

Different covariates Q to part-worths have been used in the choice modeling literature in the context of the model given by Equation (3). For example, Allenby and Ginter (1995) use demographic variables, Lenk et al. (1996) include expertise, and Chandukala et al. (2011) use consumer needs to explain variation in the part-worths. However, explaining preference heterogeneity has not met with much success generally, as manifested in performance not markedly different from a model without covariates (Rossi et al. 1996, Horsky et al. 2006).

We propose integrating a standard HB choice model and a covariate model in order to introduce dimension reduction, allow for distinct points of shrinkage, and maintain a parsimonious Γ even in the presence of a large number of covariates. Assuming covariates that are actually drivers of preference heterogeneity, a covariate model will need to accomplish three things. First, it will need to provide individual-level summaries of the covariates so we can use parameters from the covariate model as substitutes for q_n. Second, it will need to allow for each reduced dimension to be characterized by different observations. Third, it will need to be a generative model because we will rely on the results from the integrated covariate model to understand the reduced dimensions and explain preference heterogeneity. For this, a "black box" predictive model of the β_n would not be sufficient. We will evaluate the proposed and competing covariate models according to these three criteria.

2.2. Mixed Membership Models

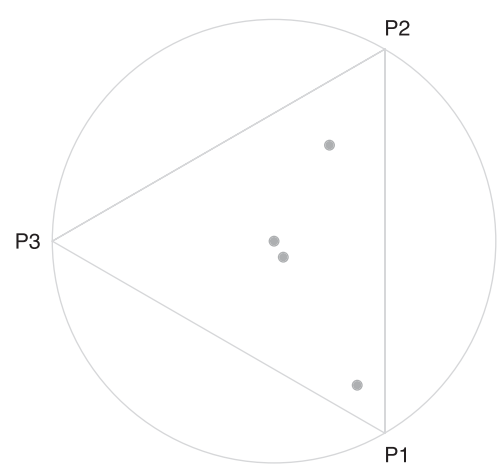
Mixed membership models are a class of models for grouping discrete data. These data can be as diverse as categorical responses in a survey to words in a document. However, unlike related models, instead of each observational unit being assigned to a particular group, a mixed membership model is characterized by having a unit-level membership vector that describes partial membership in each group. Mixed membership models also allow for different patterns in the data to be characterized by different observations.

Assume we have a collection of J discrete survey questions where each question j has L_j response categories. The probability of individual n responding to question j with response category ℓ is a function of individual n's partial membership in each of K groups as described by the K-dimensional, individual-level membership vector g_n and the response patterns for question j across K groups and L_j response categories as described by the $K \times L_j$ matrix λ_j:

$$\Pr\left(w_{n,j} = \ell|g_n,\lambda_j\right) = \sum_{k=1}^{K} g_{n,k}\lambda_{j,k}(\ell), \tag{4}$$

where K is specified by the analyst. For survey data, the K groups are known as "profiles." Each respondent is assumed to be a partial member of all profiles based on how similar their responses are to each profile. The membership vector g_n defines this individual-level mixture and is constrained so that each element is nonnegative and $\sum_{k=1}^{K} g_{n,k} = 1$. Thus, the mixed membership lives on the simplex defined by the K profiles, as illustrated in Figure 2. Profiles are often referred to as "extreme," as they define the corners of the K-dimensional simplex within which (all) observations are located. Extreme membership weights (e.g., $0,1,0,\ldots,0$) indicate identity of unit and a single profile. The $K \times L_j$ matrix λ_j is made up of K vectors $\lambda_{j,k}$ describing how likely each categorical

Figure 2. Example of Mixed Membership



Notes. This simplex is defined by three "extreme" profiles, P1, P2, and P3, located in the corners of the K space, with each of the dots representing a respondent's mixed or partial membership in each of the three profiles. Of the four cases depicted here, two are located toward one of the extreme profiles, which implies uneven mixed membership, whereas the two others have nearly equal mixed membership, which is manifested by a position at the center of the latent profile space.

response ℓ is for question j for a hypothetical respondent that is only a member of profile k. Each vector $\lambda_{j,k}$ in λ_j is also constrained with nonnegative elements so that $\sum_{\ell=1}^{L_j} \lambda_{j,k}(\ell) = 1$. Because the number of categorical responses L_j is question specific, the $J \times K \times L_j$ array $\{\lambda_j\}_{j=1}^J$ is potentially ragged.

This mixed membership model for categorical survey response data is the GoM model. It was originally developed to classify disease patterns using discrete patient-level clinical data (Woodbury et al. 1978, Clive et al. 1983) and has since been applied to modeling survey data (Erosheva et al. 2007, Gross and Manrique-Vallier 2014). However, the interest in identifying patterns across observations as well as how each observation relates to these patterns is general for all mixed membership models. What differs is the group structure and the sum-to-one constraints based on the type of data. For example, in LDA, the membership vector g_n is replaced with a topic vector θ_n that describes document n's partial membership in each topic k, and the profile array $\{\lambda_j\}_{j=1}^J$ is replaced by a topic matrix Φ that describes the weight associated with each word j for each topic k so that $\sum_{j=1}^J \Phi_k = 1$ for each k. The difference between profiles and topics is a result of the data being modeled. Respondents can repeat the same words, but they can select a response category only once for any given survey question.

The DAG in Figure 3 provides a visual representation of the GoM model. The plate notation demonstrates the three model levels: responses J, respondents N, and aggregate profiles K. To be clear, α and τ are both hyperparameters for conjugate Dirichlet priors on g_n and $\{\lambda_j\}_{j=1}^J$, where α is a vector of length K and τ is a potentially ragged array of J vectors, each with separate lengths L_j. Following Figure 3, the joint posterior distribution of the GoM model is

$$p\left(Z, G, \{\lambda_j\}_{j=1}^J | W, \alpha, \tau\right) \propto \left[\prod_{n=1}^N p\left(w_n | z_n, \{\lambda_j\}_{j=1}^J\right) p(z_n|g_n) p(g_n|\alpha)\right] p\left(\{\lambda_j\}_{j=1}^J | \tau\right), \quad (5)$$

where Z is an $N \times J$ matrix of latent profile assignments, G is an $N \times K$ matrix of membership vectors, W is an $N \times J$ matrix of observed responses, $\prod_{n=1}^N p(w_n|z_n, \{\lambda_j\}_{j=1}^J) p(z_n|g_n)$ is the likelihood, and $\prod_{n=1}^N p(g_n|\alpha)$ and $p(\{\lambda_j\}_{j=1}^J|\tau)$ are priors. If $K < J$, the model reduces the dimensions of the data from $N \times J$ to $N \times K$. In the GoM, the latent profile assignments $z_{n,j}$ are analogous to topic assignments to words in the LDA model. Both are assumed to be independent and identically distributed (i.i.d.) draws from the unit-level profiles g_n (θ_n in LDA), allowing for the collection of individual-level responses w_n to result from a mix of latent profiles (or topics). In our Bayesian estimation of the integrated choice and GoM model (Online Appendix A), the latent z_n are augmented, which allows us to "observe" how respondents switch from profile to profile as they move from question to question.

We can see from Figure 3 and Equation (5) that the GoM model satisfies our three requirements for a covariate model. First, the respondent-level membership vectors g_n can be used as substitutes for q_n in Equation (2). Second, the latent profile assignments $z_{n,j}$ allow for each observation j to potentially characterize different profiles k such that the pattern of responses defining each profile is generated only by those responses associated with the given profile. Third, the GoM is a generative model, able to explain the origin of the covariate data. Probabilities $\{\lambda_j\}_{j=1}^J$ express archetypes in a population, and the corresponding individual-level profile weights G can be used to explain response heterogeneity across respondents.

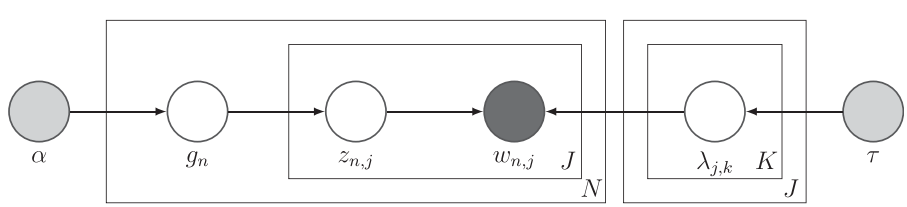
There are a number of related or competing covariate models. However, some of these do not satisfy all three requirements. We consider three competing models: a finite mixture model, a factor model, and a constrained GoM model. To illustrate model differences, the DAGs in Figure 4 provide visual representations of the GoM model and the three competing covariate models.

First, a standard finite mixture model for J discrete survey questions defines the probability of respondent n providing response ℓ to question j:

$$\Pr\left(w_{n,j} = \ell | g_j, \lambda_j\right) = \sum_{k=1}^K g_{j,k} \lambda_{j,k}(\ell), \quad (6)$$

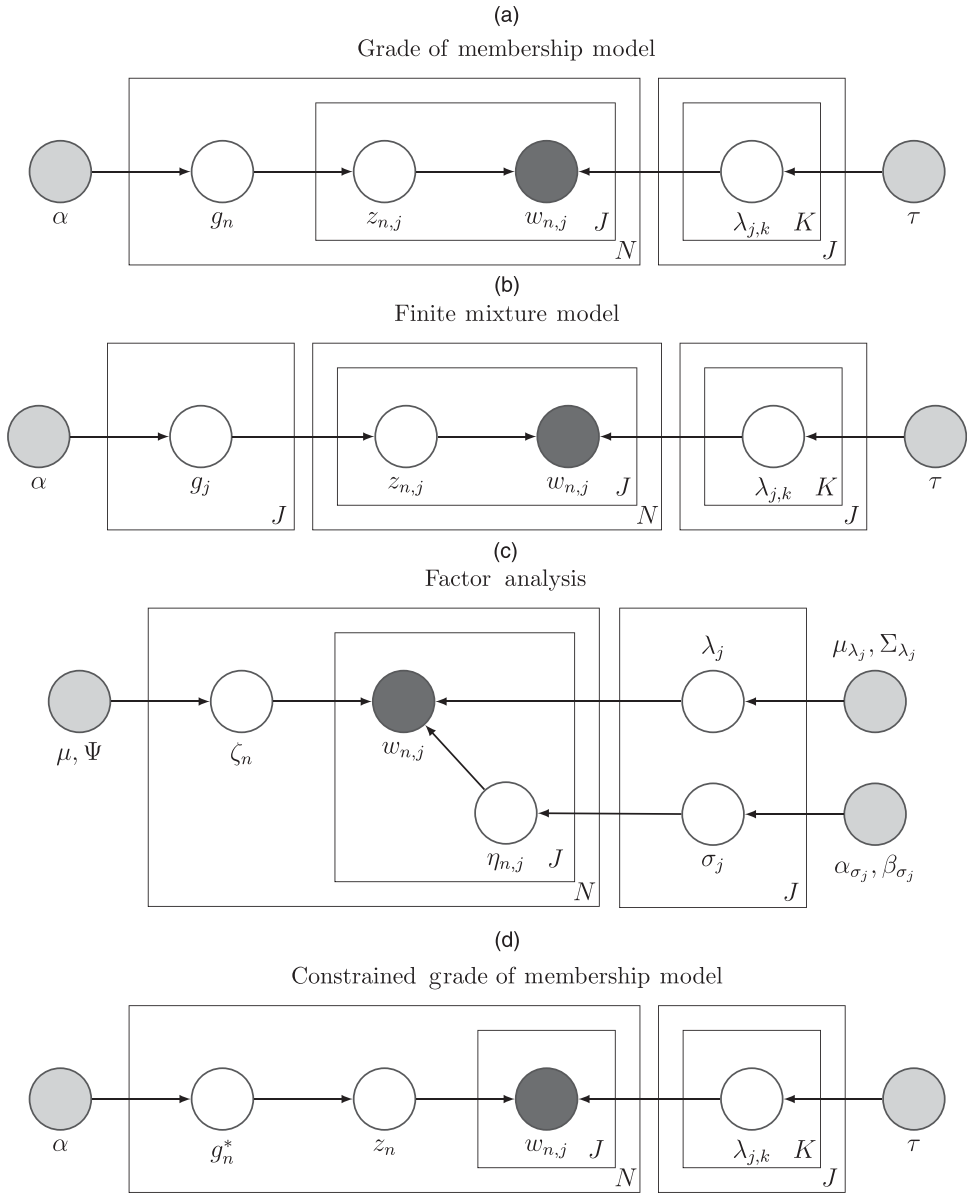
with a membership vector g_j and profiles λ_j (Kamakura and Russell 1989, Frühwirth-Schnatter 2006). Figure 4(b) is the associated DAG. It is clear that finite mixture models assume population-level membership

Figure 3. Grade of Membership Model



Note. The plates in the DAG represent replication across responses J, respondents N, and aggregate profiles K.

Figure 4. Related and Competing Covariate Models



vectors g_j and thus do not satisfy the first of our criteria for a covariate model of providing the desired summary of covariate data on the individual level. Finite mixture models can be viewed as a special case of mixed membership models (Erosheva et al. 2007, Galyardt 2014), whereas mixed membership models are often referred to as individual-level mixture models. This difference in membership vectors aligns with model use. Finite mixture models segment respondents, whereas mixed membership models describe how respondents relate to extreme profiles.

Second, factor analysis has long been a standard approach in marketing for dimension reduction (Stewart 1981). The basic assumption of factor analysis is that a set of variables can be reduced to a smaller number of latent constructs called factors. In standard factor analysis, the data for respondent n and question j are assumed to arise as follows:

$$w_{n,j} = c_j + \zeta_n' \lambda_j + \eta_{n,j}, \quad \zeta_n \sim N(0, \Psi), \quad \eta_{n,j} \sim N\left(0, \sigma_j^2\right), \tag{7}$$

where c_j is a scalar constant, ζ_n is a vector of individual-level factor scores of length K, each λ_j is a vector of regression coefficients of length K commonly known as factor loadings, and Ψ is a $K \times K$ matrix of prior covariances of ζ_n. In matrix notation,

$$w_n = c + \zeta_n' \Lambda + \eta_n, \tag{8}$$

where w_n is a vector of observations of length J, c is a common vector of variable baselines, η_n is a vector of errors of length J, and Λ is a $J \times K$ matrix of factor

loadings collected across equations. The form of factor analysis in Equation (8) is similar to that of the GoM model in Equation (4), with factor scores in place of the membership vector and factors in place of the profiles.

However, the DAG of factor analysis in Figure 4(c) reveals an important difference when compared with the GoM model: Factor analysis does not allow for each factor to be characterized by different observations (Manton et al. 1994, Marini et al. 1996). In factor analysis, the individual-level factor scores ζ_n are mapped to the covariates in a homogeneous fashion. The DAGs in Figure 4 make this clear, with λ_j specified for factor analysis and $\lambda_{j,k}$ specified for the GoM model. In other words, each regression in Equation (8) has a unique set of coefficients that hold across respondents. Thus, factor analysis satisfies only two of our three requirements.

Despite both providing dimension reduction, the goals associated with the use of factor analysis and the GoM model differ. Note that Equation (8) implies

$$w_n - c \sim N\left(0, \lambda\Psi\lambda' + I_J\sigma^2\right), \tag{9}$$

where I_J is the $J \times J$ identity matrix and σ^2 is the vector of residual variances for the variables (Lee 2007). In applying factor analysis, it is typically assumed that the factor scores are uncorrelated and the variances of the factor scores are constrained to one for identification of the scale of the factor loadings so that $\Psi = I_K$, as follows:

$$w_n - c \sim N\left(0, \lambda\lambda' + I_J\sigma^2\right). \tag{10}$$

Equation (10) demonstrates that factor analysis exploits correlation among the covariates by finding a set of factors accounting for this correlation. Thus, although factor analysis satisfies the three criteria for a covariate model, the objective of factor analysis is to uncover a limited number of latent constructs underlying a set of *variables*, whereas the objective of the GoM model is to both uncover extreme profiles of *respondents* and measure each respondent's proximity to these profiles. In other words, the GoM model has the description of respondents and respondent heterogeneity as the objects of inference.

Third, Figure 4(d) is the DAG of a constrained version of the GoM model. In this model, subjects are assumed to belong to only one of the K profiles. This is achieved by removing the z_n from the J plate. As a result, it is assumed that z_n is generated by way of a single draw from g_n^* and that all $z_{n,j}$ are identical replicates of that draw. In other words, responses for an individual are assumed to originate from $\lambda_{j,k=z_n}$. In our empirical analysis, we consider this constrained GoM model, which still satisfies the three criteria for a covariate model, to test the empirical importance of a mixed membership approach to our data. Note that this constrained GoM model is not equivalent to standard finite mixture models because g_n^* of this model resides on the N level. It is also not equivalent to factor analysis because it allows for each profile to be characterized by different observations (i.e., $\lambda_{j,k}$).

2.3. Integrated Hierarchical Bayesian Choice and GoM Model

The proposed model integrates an HB choice model and a GoM model in order to introduce dimension reduction, uncover distinct points of shrinkage for the individual-level part-worths, and maintain a parsimonious mapping from covariates to part-worth estimates (when $K << J$). By integrating these component models, we arrive at a "supervised" GoM model with choice data as additional "downstream" likelihood information to the latent G (see Figure 5). This is similar to a supervised LDA, which is prototypical of using multiple observed responses to estimate latent variables in the context of discrete data (Blei and McAuliffe 2007).

On the individual-level, our choice model remains a multinomial logit, as specified in Equation (1), and the distribution of heterogeneity remains multivariate normal as in Equation (2). Because the GoM provides g_n for each respondent, as shown in Equation (4), we use these membership vectors in the place of the covariates q_n to explain heterogeneity in the part-worths B (i.e., $\beta_n = \Gamma g_n + \xi_n$). The likelihood of the integrated model is

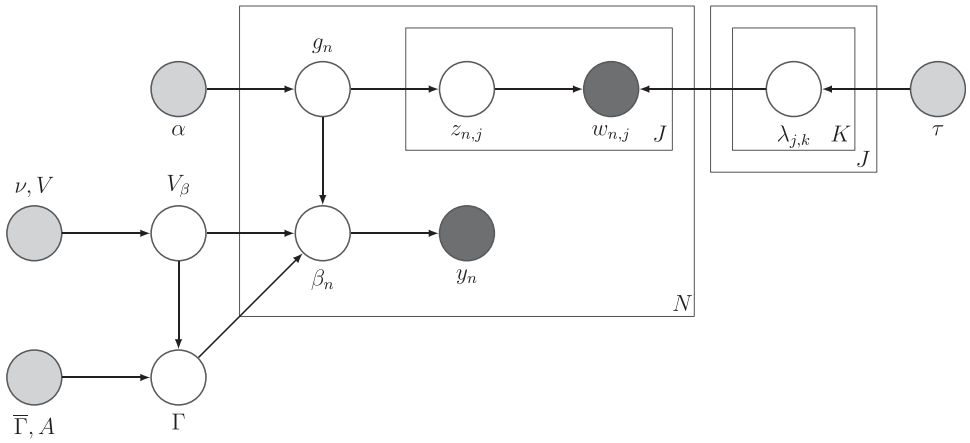
$$p\left(Y, W|B, \Gamma, V_\beta, Z, G, \{\lambda_j\}_{j=1}^J\right) = \prod_{n=1}^N p(y_n|\beta_n)p(\beta_n|g_n, \Gamma, V_\beta)p\left(w_n|z_n, \{\lambda_j\}_{j=1}^J\right)p(z_n|g_n). \tag{11}$$

Figure 5 illustrates the proposed integrated HB choice and GoM model. From the DAG, we can see that the proposed model is a three-level model where only the categorical response vector w_n and choice vector y_n for each respondent are observed. The membership vector g_n serves as the link between the choice and GoM components of the model. Thus, g_n is informed by both the categorical responses w_n and the chosen alternatives y_n.

Following Figure 5, the joint posterior distribution of the proposed model is

$$p\left(B, \Gamma, V_\beta, Z, G, \{\lambda_j\}_{j=1}^J|Y, \bar{\Gamma}, A, \nu, V, W, \alpha, \tau\right) \propto \left[\prod_{n=1}^N p(y_n|\beta_n)p(\beta_n|g_n, \Gamma, V_\beta)p\left(w_n|z_n, \{\lambda_j\}_{j=1}^J\right) \cdot p(z_n|g_n)p(g_n|\alpha)\right]p(\Gamma|V_\beta, \bar{\Gamma}, A)p(V_\beta|\nu, V)p\left(\{\lambda_j\}_{j=1}^J|\tau\right), \tag{12}$$

Figure 5. Integrated Hierarchical Bayesian Choice and GoM Model



Notes. From the DAG, we can see that the membership vector g_n serves as the link between the choice and GoM components of the model. The membership vector g_n is informed by both the categorical responses w_n and the chosen alternatives y_n.

where $\prod_{n=1}^{N} p(y_n|\beta_n)p(\beta_n|g_n, \Gamma, V_\beta)p(w_n|z_n, \{\lambda_j\}_{j=1}^{J})p(z_n|g_n)$ is the likelihood, and $\prod_{n=1}^{N} p(g_n|\alpha), p(\Gamma|V_\beta, \bar{\Gamma}, A), p(V_\beta|\nu, V)$, and $p(\{\lambda_j\}_{j=1}^{J}|\tau)$ are the priors. A complete list of the variables in Equation (12) is given in Table 1. Details on estimation, model validation, and computational complexity are provided in Online Appendix A.

In our model, Γ maps complete sets of individual-level profiles weights g_n to part-worths β_n; that is, for any admissible combination of $g_{n,k}$, Γ generates a prior expectation with respect to β_n. Given the constraint to profile weights in the GoM model (i.e., $\sum_{k=1}^{K} g_{n,k} = 1$), any combination of $g_{n,k}$ with nonnegative values that add up to 1 is admissible, and Γ can always be interpreted as providing their mapping to β_n. However, as regression coefficients, Γ cannot be interpreted in the standard way (i.e., change of y as x_1 changes by one unit, holding all other x fixed). This is because any $g_{n,k}$ cannot change independently from $g_{n,-k}$; as one weight in g_n changes, at least one other weight must change as well.

If a standard interpretation of Γ is desired, one needs to transform the compositional covariates g_n. One way to achieve this is by applying the isometric log-ratio (ILR) transformation with respect to profile K, where the ordering of profiles is not consequential:

$$g_{n,k}^{\dagger} = \sqrt{\frac{K-k}{K-k+1}} \times \ln \frac{g_{n,k}}{\sqrt[K-k]{\prod_{i=k+1}^{K} g_{n,i}}}, \tag{13}$$

Table 1. Variables in the Integrated Hierarchical Bayesian Choice and GoM Model

Variables	Description
Choice variables	
N	Number of respondents
H	Number of choice tasks for each respondent n
P	Number of alternatives in each choice task
M	Number of attribute levels in each choice task
y_n	H-dimensional vector of choices for respondent n
β_n	M-dimensional vector of part-worths for respondent n
Γ	$K \times M$ matrix of regression coefficients mapping individual-level profiles weights g_n to part-worths β_n
V_β	$M \times M$ covariance matrix of the random effects distribution of heterogeneity
GoM variables	
K	Number of profiles
J	Number of categorical questions
L_j	Number of categorical responses for question j
w_n	J-dimensional vector of respondent n's categorical responses
z_n	J-dimensional vector of respondent n's profile assignments
g_n	K-dimensional membership vector for respondent n
$\{\lambda_j\}_{j=1}^{J}$	Array of probability distributions $\lambda_{j,k}$ over the L_j response options for each question j and profile k

which projects g_n onto an orthonormal base. This orthonormal base allows for a standard interpretation of Γ as the $g^{\dagger}_{n,k}$ are unconstrained and can therefore move independently. Note that if $g_n = (1/K, \ldots, 1/K)$, the resulting $g^{\dagger}_n = (0, \ldots, 0)$, which implies that the baseline of a regression model using ILR-transformed compositional data indicates the expected value of the dependent variable when all original component shares are equal (i.e., centered within the simplex as in Figure 2). Also note that as result of the transformation in Equation (13), $g^{\dagger}_{n,K} = 0$. In the following, we use $g^{\dagger}_n$ to describe the transformed membership vectors, omitting the values from the Kth profile, as this carries no information, and adding a baseline so that $g^{\dagger}_n = (1, g^{\dagger}_{n,1}, g^{\dagger}_{n,2}, \ldots, g^{\dagger}_{n,K-1})$. In the likelihood of our model in Equation (11), the ILR transformation itself can actually be ignored because Equation (13) is deterministic, which implies $p(g^{\dagger}_n | g_n) = 1$. In our empirical analysis, we use the model with ILR-transformed G for obtaining estimates of cross-sectional Γ so that its coefficients can be interpreted the usual way. We provide details on estimating the integrated choice and GoM model using ILR-transformed G in Online Appendix A.

3. Boundary Conditions

In this section, we present results from a simulation study in which we explore the boundary conditions of our proposed integrated choice and GoM model. More specifically, we investigate the conditions under which it is possible to empirically distinguish the proposed model from a set of competing models when the data are generated by the proposed model. Each of the models under consideration differs in terms of the nature of the regressors used in the upper-level of the choice model. The models are the proposed integrated choice and GoM model with the membership vector as regressors (i.e., the membership vector model), the standard model with observed binary covariates as regressors (i.e., the binary covariate model), the integrated choice and factor analysis model with the factor scores as regressors (i.e., the factor score model), and the integrated choice and constrained GoM model (i.e., the constrained membership model). The constrained membership model arises from assuming that all profile assignments $z_{n,j}$ arise from a single draw from g_n, as detailed in Section 2. The binary covariate, factor score, and constrained membership models are included in the simulation study because they are the standard approach to using covariates or the primary competing covariate models, respectively.

We evaluate performance by computing the ability of the model to predict the complete vector of part-worth utilities for holdout respondents. To achieve this, we use in-sample data to conduct inference with respect to how observed covariates map to the part-worths and then predict these part-worths for out-of-sample respondents using the mapping scheme implied by each upper-level model. Note that we do not use holdout tasks for this analysis or likelihood-based measures of in-sample fit. Likelihood-based fit measures or accuracy of predicting holdout tasks implies choices on the individual level to be observed. The upper-level model plays only a partial role in making such calculations and will be influential only if the number of observed choices is small. Our goal is different. We want to answer the question of how well we can predict the distribution of heterogeneity for a new sample of respondents. For this, we compute the hit probability across all choices of holdout respondents given knowledge of the mapping of covariates to part-worths implied by the upper level of each model.

For this analysis, we generate 200 data sets from the membership vector model. For each data set, we fix $N = 500$, $L_j = 2$ for all j, and $V_\beta = 0.33 I_M$ throughout, and generate $\Gamma_{j,m} \sim U(-4, 4)$ for all j and m. Our search grid to establish boundary conditions is defined by the following variables, for which we randomly pick a single value for each data set from the indicated range:

- $J \in (5, 6, \ldots, 20)$,
- $K \in (2, 3, \ldots, 8)$, and
- $\alpha, \tau \in (0.1, 0.5, 1, 2, 4, 8, 16)$.

The number of responses generated ($J \leq 20$) is similar to what most marketing surveys, given respondents' time constraints, contain. The number of latent archetypes (K) is, on average, smaller than J, reflecting the typical desire in marketing research to reduce dimensionality and complexity in describing heterogeneity. In applications of latent class models, managers are often limited to a small number of types; α and τ are generated from a large range of values, allowing for very different relationships between extreme types and observables and also different distributions of weights in the population. Having obtained draws of these variables, we generate G and $\{\lambda_j\}_{j=1}^J$ from Dirichlet distributions, given values of α and τ, respectively. Given G and $\{\lambda_j\}_{j=1}^J$, we generate J binary covariates (because $L_j = 2$ for all j) for each respondent through draws from a binomial distribution. The draw of β_n given G, $\{\lambda_j\}_{j=1}^J$, and V_β is then $\beta_n \sim \mathrm{MVN}(\Gamma g_n, V_\beta)$. The simulation of choice data then proceeds in the usual way. We generate $P = 4$ choice options for each of $H = 15$ tasks, each consisting of $M = 12$ attribute levels. For each choice set, the "observed" choice is given by the option with highest utility, given $X\beta_n + \xi$, where ξ is a vector of P independent error draws from an extreme value distribution. We generate entries of the $P \times M$ design matrix of choice options X i.i.d. from $N(0, 2)$.

The search grid defined by our setup considers different situations. Small values for α characterize a situation in which respondents are similar to a single

extreme while large values for α characterize a situation in which respondents have more of a mixed membership across profiles. Small values for τ indicate that, given a profile k, only few covariates are indicative of that profile. The term K specifies the dimensionality of the underlying latent profile space. Note that we consider scenarios where $K > J$, implying that there are more extreme profiles than measurement indicators in the population. Our goal is to establish the conditions under which the true data-generating mechanism performs better than competing models.

The out-of-sample hit probability is the average posterior probability of a set of holdout respondents N^* and post-burn-in Markov chain Monte Carlo (MCMC) draws R given a model D:

$$\mathrm{HP}(D) = \frac{1}{N^*}\sum_{n^*=1}^{N^*}\left[\frac{1}{H}\sum_{h=1}^{H}\left(\frac{1}{R}\sum_{r=1}^{R}\Pr\left(y_{n^*,h}|\beta_{n^*,r}^D, X_h\right)_{n^*}\right)\right], \quad (14)$$

where $y_{n^*,h}$ is the observed choice for holdout respondent n^* for choice task h, and $\beta_{n^*,r}^D$ are respondent n^*'s estimated coefficients for post-burn-in MCMC draw r and model D. These $\beta_{n^*,r}^D$ are drawn from the distribution of heterogeneity $\mathrm{MVN}(\Gamma_r^D q_n^D, V_{\beta,r}^D)$ for the binary covariate model.

For the proposed membership vector model, we need to first generate out-of-sample membership vectors $g_{n^*}^D$ before we can draw $\beta_{n^*,r}^D$ from the distribution of heterogeneity $MVN(\Gamma_r^D g_{n^*}^D, V_{\beta,r}^D)$. For this, we first generate an initial $g_{n^*}^D$ from Dirichlet(α^*). In the absence of a model that estimates α from calibration data, we determine α^* by minimizing the Kullback–Leibler divergence of the "observed" distribution of g_n^D and Dirichlet(α^*) with respect to possible values of α^*. The membership vectors $g_{n^*,r}^D$ are then successively updated for each of the post-burn-in MCMC draws using the out-of-sample sample response data $w_{n^*,j}$, the draws of $\lambda_{j,k,r}^D$, and α^*:

$$\begin{aligned} z_{n^*,j,r}^D &= \mathrm{Multinomial}\left(p_1^D, \ldots, p_K^D\right), \text{ where} \\ p_k^D &\propto g_{n^*,k,r}^D \lambda_{j,k,r}^D\left(w_{n^*,j}\right), \\ g_{n^*,r}^D &\sim \mathrm{Dirichlet}\left(z_{n^*,j,r}^D + \alpha^*\right). \end{aligned}$$

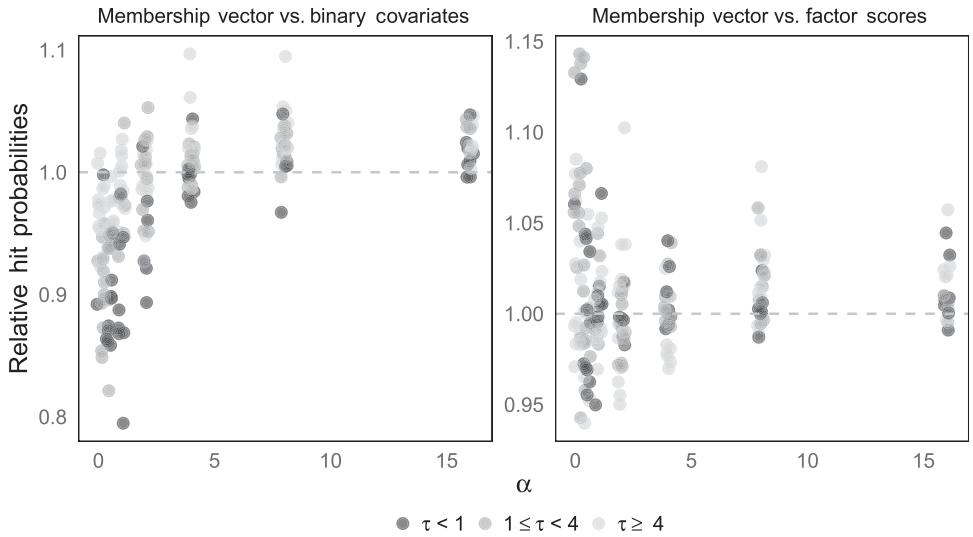
We can then draw $\beta_{n^*,r}^D \sim \mathrm{MVN}(\Gamma_r^D g_{n^*,r}^D, V_{\beta,r}^D)$ for the out-of-sample respondents and compute the hit probability. A similar process is followed for the constrained membership model. To compute hit probabilities for the factor score model, we first generate draws of ζ_n^D using the observed covariates and across-subject factor loadings Λ. For this step, we use standard conjugate results from Bayesian factor modeling (Lee 2007). Given the ζ^D, we generate $\beta_{n^*}^D$ using results from the hierarchical regression by way of drawing from $\mathrm{MVN}\left(\Gamma'\zeta_{n^*}^D, V_\beta\right)$.

To compute hit probabilities for out-of-sample respondents, we use data from $N = 250$ randomly selected respondents to calibrate the model and then predict all $H = 15$ choices for the $N^* = 250$ out-of-sample respondents. We compare the binary covariate model, the factor score model, and the membership vector model with respect to this measure. A first result from our simulations is the average hit probability obtained from alternative models. These probabilities are 50.4%, 51.0% and 49.3%, averaged over the 200 data sets, for the membership vector, binary covariate, and factor score models, respectively. This suggests that the proposed membership vector model is not generally preferred in terms of fit to out-of-sample respondents across all (and equally weighted) scenarios considered here. The implication is that our *set* of simulated data sets is not skewed toward a particular model a priori. On average, the binary covariate model seems to better predict out-of-sample part-worths and implied choices. In Figure 6, we show how relative hit probabilities vary with values of α and τ, marginalized with respect to K and J.

It is clear from Figure 6 that the relative performance of the true model is dependent on how the data are generated. Relative to the binary covariate model (left panel), performance of the membership vector improves as both α and τ get larger. For small values of these parameters (0.1, 0.5), the performance disadvantage of the membership vector model can be as high as 15%. Small values for α and τ imply that respondents are well characterized by a single extreme profile of which a small set of covariates is highly indicative (probabilities approaching 1). Under such conditions, unobserved membership to latent profiles can be substituted by observed covariates. If, however, α and τ get large (>2), this is no longer possible, and the performance advantage of the membership vector over the binary covariate model can exceed 10% (Figure 6). In other words, if respondents occupy the corners of the latent K space and single covariates describe this position well, there is little to be gained by the membership vector model compared with the binary covariate model. Both will result in very similar descriptions of the distribution of heterogeneity of preferences in the data. This is no longer true when respondents are a mix of profiles (i.e., $\alpha \gg 1$) and covariates do not exhibit a deterministic relationship to profiles (i.e., $\tau \gg 1$). The right panel of Figure 6 shows the hit probabilities of the membership vector model relative to the factor score model, revealing a reversed pattern: Larger values for α and τ lead to the factor model generating hit probabilities almost equal to the membership vector model. Instead, for the membership vector model to perform better, τ needs to be small.

For a more granular analysis of results from the simulation, we employ a logit regression model using all parameters in our simulation simultaneously (Table 2).

Figure 6. Relative Out-of-Sample Hit Probabilities



Notes. The figure shows jittered relative out-of-sample hit probabilities given a range of values for α and τ for the membership vector and binary covariate models (left panel) and the membership vector and factor score model (right panel). Each dotted grey line is at a relative predictive fit of 1, meaning both models perform the same. Values above 1 indicate the membership vector performs better relative to the competing model, and values below 1 indicate the opposite.

The dependent variable in this regression analysis is a binary indicator of the true membership vector model outperforming the benchmark model with respect to out-of-sample choice predictions. That is, if the average hit probability from the membership vector across all 250×15 out-of-sample choices is higher than that of the benchmark model considered, we code this result as 1 and otherwise as 0. As covariates, we include the main effects of our simulation parameters as well as all two-way interaction effects. We find a (binary) logit regression model to be a more accurate summary of our simulation results than a linear regression model (and using the relative hit probabilities as the dependent variable) because of the highly nonnormal distribution of residuals from the linear model and the nonlinear relationship of many covariates to the indicator of model preference (see Figure 6). As benchmark models to the membership vector model, we consider the binary covariate model, the factor score model, and the constrained membership model with individual-level profiles weights constrained to be 0 or 1.

Table 2. Performance of Membership Vector Model Relative to Alternative Models

	Binary covariates		Factor scores		Constrained membership	
	Coefficient	SE	Coefficient	SE	Coefficient	SE
Intercept	1.747	2.039	−2.069	1.600	−3.451	2.729
α	0.450	0.431	−0.191	0.215	**−1.917**	0.549
τ	−0.176	0.201	−0.210	0.153	−0.010	0.180
K	−0.653	0.366	0.232	0.288	0.261	0.465
J	−0.233	0.155	0.151	0.106	0.154	0.174
$\alpha \times \tau$	**0.153**	0.049	0.004	0.012	−0.002	0.026
$\alpha \times K$	0.072	0.042	**0.086**	0.032	**0.528**	0.130
$\alpha \times J$	−0.029	0.024	−0.007	0.011	−0.007	0.019
$\tau \times K$	**0.071**	0.030	0.034	0.019	0.025	0.023
$\tau \times J$	−0.031	0.016	0.000	0.007	−0.003	0.009
$K \times J$	0.038	0.026	−0.014	0.018	−0.024	0.029
Hit rate	86%		70%		87%	

Notes. Reported are maximum likelihood estimates from a binary logit model with preference for the membership vector model over the comparison benchmark model (i.e., yes or no) as the dependent variable. Coefficients credibly different from 0 (95% level or higher) are in bold. Hit rate indicates share of correct predictions. SE, Standard error of estimate.

Table 2 reveals that the interaction effects of our simulation parameters are important drivers of the membership vector being distinguishable from the benchmark models. Relative to the binary covariate model, the membership vector model is preferred when both α and τ increase, confirming results from our marginal analysis (Figure 6). We also find that as both α and K increase, the membership vector is preferred over both the binary covariate model and the factor score model, as indicated by positive coefficients in both analyses. For the binary covariate model, the $\beta^{(reg)}$ of covariate $\alpha \times K$ is 0.07 (probability to be different from 0 is 92.8%), and for the factor score model it is 0.09 (probability to be different from 0 is 99.3%). We obtain a similar result from covariate $\tau \times K$. As this variable increases, the probability of the membership vector model being preferred over both the binary covariate and factor score models increases. We note that the estimate of this coefficient for the factor score model (0.034) has a probability of 92.3% of being different from 0. We find the main

effect of K as well as that of J to be effectively 0 for all models, and we also find that $K > J$ is not a scenario favoring any model, implying that the dimensionality of the latent space plays no role as such. With respect to the constrained membership model, we find regression coefficients from our analysis to be very similar to those in the factor score model. This suggests that the conditions that allow us to distinguish the membership vector model from the factor score model also allow us to distinguish it from the constrained membership model. We also note that the model-based results in Table 2 allow for prediction of the relative performance of our models outside the search grid of our simulation (e.g., $J > 20$).

In conclusion, we find the proposed membership vector model to be identifiable as the true model compared with both the binary covariate model and the factor score model when respondents exhibit partial membership to multiple extreme profiles (i.e., larger α) and, simultaneously, the relationship of profiles to the observed covariates is of a probabilistic, not deterministic, nature (i.e., larger τ). The latter is the case when the probability of observing a certain outcome, given profile k, approaches 1. The analysis also shows that the number of latent profiles K plays an important role, but only in concert with α or τ. A larger number of latent profiles works in favor of the membership vector model only if α or τ also increase. This suggests that the membership vector model does not benefit from more extreme types in the population as such. These types must contribute to partial membership (i.e., not push respondents into corners of the K space) and not be uniquely measurable by specific covariates.

An implication of this result is that the standard approach in marketing (using covariates directly as predictors of part-worths) works reasonably well when respondents (a) are effectively one of the latent types and (b) these types can be identified by specific covariates. Some combination of the observed covariates is then a reasonably good predictor of membership to latent types, implying that this observed pattern can approximate the latent g_n well. The factor score model, a typical approach to latent variable modeling in marketing, works well under similar conditions, combined with a relatively small number of latent types. Neither the factor score model nor the covariate model is an appropriate modeling approach, however, if respondents truly exhibit mixed membership (i.e., g_n away from 0 and 1) and if the number of latent types K is large.

4. Empirical Applications

In our empirical analysis, we use data from two surveys of preferences regarding robotic vacuums and smartphones using national samples in the United States and Germany, respectively. Whereas the robotic vacuum data come from an emerging market, the smartphone data represent a well-established market, providing us a broad test for the proposed membership vector model. For the robotic vacuum data, a total of 332 respondents were carefully screened to ensure that the product options under consideration were relevant to them. In particular, qualified respondents had to own a robotic vacuum, currently be shopping for their first robotic vacuum, or might consider a robotic vacuum sometime in the next five years. For the smartphone data, a total of 147 respondents were similarly screened to ensure they were in the market for a new smartphone.

Before the conjoint experiment, respondents were asked to detail why the product was relevant to them or anyone in their household. For the robotic vacuum data, respondents selected from a list of 11 statements on cleaning that robotic vacuums might help address and a list of 7 statements that described problems with robotic vacuums. The combined list of 18 statements regarding cleaning and robotic vacuums is provided in Table 3. For the smartphone data, respondents selected from a list of 53 statements on smartphones that described their interests and usage. A subset of the 53 statements regarding smartphones is provided in Table 4. The full list is given in Online Appendix D. Thus, our discrete data consist of two possible response categories (i.e., $L_j = 2$) for all $J = 18$ or $J = 53$ where not selecting an item is coded as 0 and selecting an item is coded as 1.

After selecting from applicable statements on cleaning and robotic vacuums or on smartphone interest and usage, respondents proceeded through a series of choice tasks where they were asked to select which of a given number of product alternatives they most preferred. For the robotic vacuum data, this set of alternatives included an outside option to not select

Table 3. Statements on Cleaning and Robotic Vacuums

No.	Item
1.	I enjoy coming home to a clean house.
2.	I don't feel relaxed when I know my home isn't clean.
3.	I worry about pet hair and dander in the home.
4.	I have trouble keeping the floor beneath my furniture clean.
5.	I worry about germs and dirt on my floor and carpet.
⋮	⋮
14.	Robotic vacuums often need to be "rescued" because they get stuck.
15.	Robotic vacuums need to have their trash containers changed too often.
16.	Robotic vacuums don't do a good enough job cleaning the floor and carpet.
17.	Robotic vacuums don't spend enough time on really dirty spots on the floor.
18.	Robotic vacuums scare household pets.

Table 4. Statements on Smartphone Interest and Usage

No.	Item
1.	The security of the OS on my SP is very important to me.
2.	The apps on my SP only run with the newest OS.
3.	A more recent OS is worth a higher price.
4.	A more recent OS shows that a SP is up-to-date.
5.	It's always useful to have a more recent OS.
6.	I don't care about the OS on my SP.
7.	I want to use my SP to make payments.
⋮	⋮
49.	During the night, I always switch my SP off.
50.	I switch my SP to vibrate in the night.
51.	I switch my SP to vibrate only when necessary.
52.	I like to surf the web on my SP.
53.	I use QR codes with my SP.

Note. OS, operation system; SP, smartphone.

any of the given alternatives. For smartphones, no such option was given. Each alternative was composed of separate attributes. Figure 7 is a screenshot of one of these choice tasks from the robotic vacuum data. The estimable attribute levels, excluding the reference levels in grey, are included in Tables 5 and 6. Note that the smartphone data do not include an outside option.

For the robotic vacuum data, we see from Table 5 that the attributes are defined in terms of brand, price, and different features, including the vacuum's performance (i.e., what percentage of dirt and debris it picks up), capacity (i.e., how often it needs to be emptied), the type of navigation (i.e., does it change directions by just bumping into things or is it "smart" and able to scan and determine an optimal path), where it can be programmed, and whether virtual borders can be set to keep the robotic vacuum away from certain areas of the home. For the smartphone data, we see from Table 6 that the attributes are defined in terms of functional attributes, including display size (i.e., the effective size of the phone), display resolution, camera quality for the front camera, available memory, and price. A summary of the data sets using model notation is provided in Table 7.

5. Results

In this section, we report results from applying the proposed membership vector model to our data sets. We start by comparing the fit of the proposed model to several competing models that differ in terms of their treatment of covariates. We then present results from the proposed membership vector model.

5.1. Model Comparison

For comparison of our proposed model, we consider several alternative upper-level models for covariates. A natural benchmark is the binary covariate model, which uses the observed q_n as (exogeneous) regressors, representing the standard practice of using these variables in a choice model. The proposed membership vector model uses the membership vectors g_n from the integrated GoM model as regressors. The factor score model utilizes an alternative assumption about the latent structure of the covariates and uses the factor scores ζ_n from the integrated factor analysis model as regressors. Estimation details for this model are provided in Online Appendix B. The constrained membership vector model assumes respondents belong to only one of the K profiles and uses the (constrained) membership vector g_n^* as regressors. The factor scores and constrained membership vector models are the primary competing covariate models.

If the binary covariate model is the standard approach to using covariates and the membership vector, factor score, and constrained membership vector models represent competing ways to model covariates, the last alternative model considered provides a third approach. We use a Bayesian variable selection (VS) model operating on main and interaction effects of the observed covariates. The VS model applied here is a multivariate version of the stochastic search VS

Figure 7. (Color online) Example Robotic Vacuum Choice Task

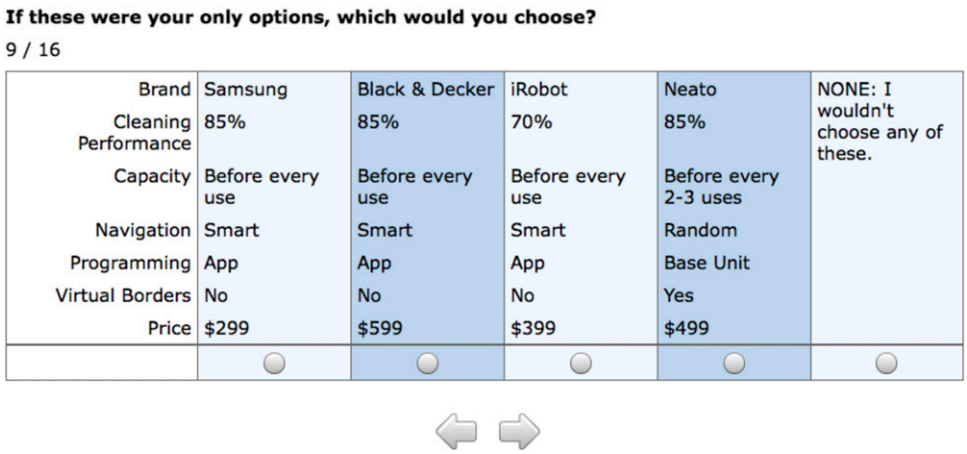


Table 5. Robotic Vacuum Attribute Levels

Attributes	Levels				
Brand	Outside option	Neato	iRobot	Samsung	Black & Decker
Performance	70%	85%			
Capacity	Every use	Every 2–3 uses			
Navigation	Random	Smart			
Programming	Base unit	App			
Virtual borders	No	Yes			
Price	$299	$399	$499	$599	

model proposed by George and McCulloch (1993). The upper-level model for individual n is then $\beta_n = \Gamma(\delta q_n) + \xi_n$, where δ is a cross-sectional vector of binary latent variables indicating inclusion of the (complete) covariate vector q_n to the design matrix. Details for this model are provided in Online Appendix C. In our application, we consider all two-way and all three-way interactions for the robotic vacuum data and all two-way interactions for the smartphone data, giving rise to $J = 1{,}159$ and $J = 1{,}432$ covariates for the robotic vacuum data and the smartphone data, respectively. We include interactions for the VS model to allow for a large search space δ so that the VS model can empirically differ from the binary covariate model, which is not necessarily the case when only main effects are considered. The VS model can, of course, result in the binary covariate model (i.e., when $\delta = 0$ for all interactions). Estimation of the VS model using our data results in 95% and 77% of covariates being excluded for the smartphone data and robotic vacuum data, respectively. This implies that coefficients are credibly different from zero for a selection of interaction variables in both data sets. In our model comparison, we show that this, however, does not favor the VS model.

We report five fit statistics for each of the models on each of the data sets: three in-sample and two predictive. Prior to computing any fit statistic, we checked for but found no substantial evidence of label switching. Fit statistics reported for the membership vector model are based on the GoM model without ILR transformation of G. Our first in-sample fit measure is the Newton and Raftery (1994) approximation of the log marginal density (LMD), a standard Bayesian measure for model fit. The second is the deviance information criterion (DIC; Spiegelhalter et al. 2002), which explicitly penalizes overparameterized models. The third is a Bayesian-specific improvement on the DIC, the Watanabe–Akaike information criterion (WAIC; Watanabe 2010). For all three in-sample fit statistics, values closer to zero indicate improvement in fit. We only consider the likelihood contribution from the observed choices for purposes of model comparison, which allows to directly compare standard and integrated models. Consider that the binary covariate model treats covariates as exogenous, which implies that these provide no likelihood contribution. For computation of out-of-sample predictive fit, we use hit rates (i.e., correct prediction, yes or no) and hit probabilities as detailed in Section 3.

Table 6. Smartphone Attribute Levels

Attribute	Levels			
Display size	4 inch	4.7 inch	5 inch	5.5 inch
Display resolution	Standard	HD		
Camera (front)	4 MP	6 MP	8 MP	12 MP
Memory	4 GB	16 GB	32 GB	64 GB
Price	$400	$600	$800	

Note that in our application of latent variable models, the number of profiles or factors was not part of the model(s) and determined by us. Joutard et al. (2007) provides a review of the various model selection criteria for mixed membership models, from using a Dirichlet process prior to estimate K to using various information criterion to select K. Given the added computational complexity of estimating K, we chose to use LMD and the DIC to select K. For the membership vector, factor scores, and constrained membership models, we first ran the covariate models on covariates in Tables 3 and 4 and then determined the number of profiles or factors given fit.

As an example of our approach, Figure 8 shows LMD and the DIC for GoM models given $K = 2$ to $K = 18$ run on the robotic vacuum data. In terms of fit, both LMD and the DIC indicate a solution with a (relative to $J = 18$) medium number of profiles (five to seven). Whereas this procedure is helpful, we have also found an evaluation of the profiles themselves to be informative when choosing K. With the range of possible models narrowed, we ran the proposed membership vector model for $K = 5$ to $K = 7$. Comparing results to find profiles that are sufficiently differentiated and nonrepeating, the model with $K = 5$ was deemed best. Determining differentiated profiles is a somewhat subjective step but is essential in order to find sufficiently different archetypes to define the simplex in the latent K space. The same process was repeated for the smartphone data. A similar procedure was followed

Table 7. Data Summary

Variables	Robotic vacuums	Smartphones
N, total number of respondents	332	147
H, number of choice tasks for each respondent n	16	17
P, number of alternatives in each choice task	5	3
M, number of attribute levels in each choice task	12	12
J, number of categorical questions	18	53
L_j, number of categorical responses for each question j	2	2

to select K for the factor scores and constrained membership models.

Table 8 presents in-sample and out-of-sample fits across all models and data sets. The model exhibiting the best fit, given measure of fit, is indicated in bold. Predictive fit is computed on a random holdout sample of 20% of respondents from each data set. With respect to in-sample fit, we find that the variable selection model fits the data best for both data sets. Recall that this model operates on main and (included) interaction effects. This allows for very flexible adaption of this model to the data and, hence, excellent in-sample fit. Putting aside the VS model as an outlier in terms of in-sample fit, the proposed membership vector model performs best among the remaining models, as indicated by italics in the table, for all of the in-sample fit statistics for the smartphone data set. That said, it does not perform markedly differently from either the binary covariate model or other latent variable models for either data set, suggesting that for estimation of β_n from choice data, specification of the upper level of the model plays a minor role in terms of in-sample fit.

Out-of-sample predictive fit uniformly favors the proposed membership vector model. Compared with the second-best-fitting integrated model (for smartphones, the factor score model, and for robotic vacuums, the constrained membership model), this model results in lifts of hit rates of 15% and 12% for robotic vacuum and smartphone data, respectively. Compared with the binary covariate model, the lift is 15% for robotic vacuums and 19% for smartphones. Relative to the flexible variable selection model, the proposed membership vector model lifts hit rates by 33% for robotic vacuums and 9% for smartphones. Because the purpose of our model is explaining preference heterogeneity, predicting all the choices of holdout respondents is the best indicator of model performance. Given our empirical results, we therefore conclude that a mixed membership approach to modeling covariates in a choice model is the best way to capture the true distribution of preferences.

The variable selection model, which performs so well in-sample, is outperformed with respect to out-of-sample predictive fit by both the membership vector model and the constrained membership vector model for both data sets, suggesting that a direct approach to explaining part-worths is not preferred even when (many) interactions are potential drivers.

Figure 8. Selecting K for the Robotic Vacuum Data Set

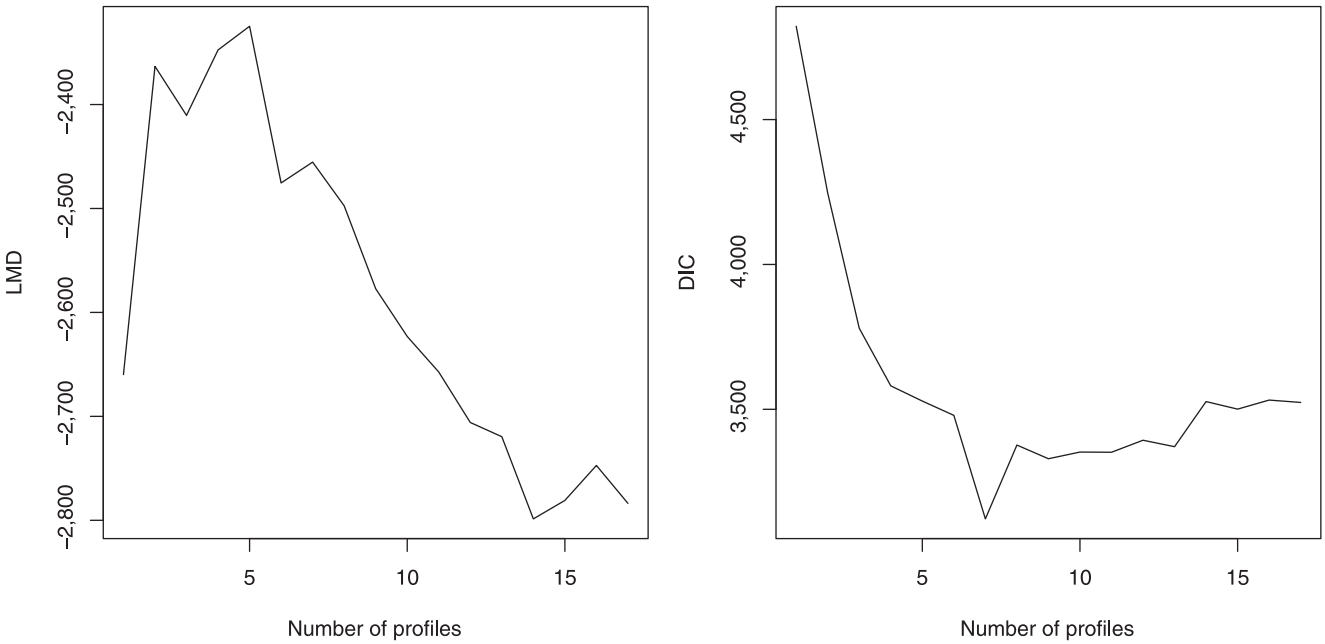


Table 8. Model Fit

	In-sample			Predictive	
Models	LMD	DIC	WAIC	Hit rate	Hit prob.
Robotic vacuum data					
Binary covariates ($\beta_n = \Gamma q_n + \xi_n$)	–1,904	7,137	6,182	0.401	0.392
Membership vector ($\beta_n = \Gamma g_n + \xi_n$)	–1,806	6,842	*5,349*	**0.462**	**0.443**
Factor scores ($\beta_n = \Gamma \zeta_n + \xi_n$)	–1,848	6,911	5,653	0.347	0.327
Constrained membership ($\beta_n = \Gamma g_n^* + \xi_n$)	*–1,794*	*6,732*	6,528	0.400	0.400
Variable selection ($\beta_n = \Gamma(\delta q_n) + \xi_n$)	**–1,521**	**5,717**	**4,394**	0.347	0.342
Smartphone data					
Binary covariates ($\beta_n = \Gamma q_n + \xi_n$)	–859	3,163	2,850	0.453	0.449
Membership vector ($\beta_n = \Gamma g_n + \xi_n$)	*–744*	*2,687*	*2,314*	**0.540**	**0.526**
Factor scores ($\beta_n = \Gamma \zeta_n + \xi_n$)	–905	3,360	3,030	0.480	0.482
Constrained membership ($\beta_n = \Gamma g_n^* + \xi_n$)	–888	3,334	2,372	0.440	0.443
Variable selection ($\beta_n = \Gamma(\delta q_n) + \xi_n$)	**–607**	**2,171**	**1,863**	0.494	0.483

Note. The best-performing model for the given data set and fit statistic is in bold. Similarly, the second best-performing model for the given data set and fit statistic (in-sample only) is in italics; prob, probability.

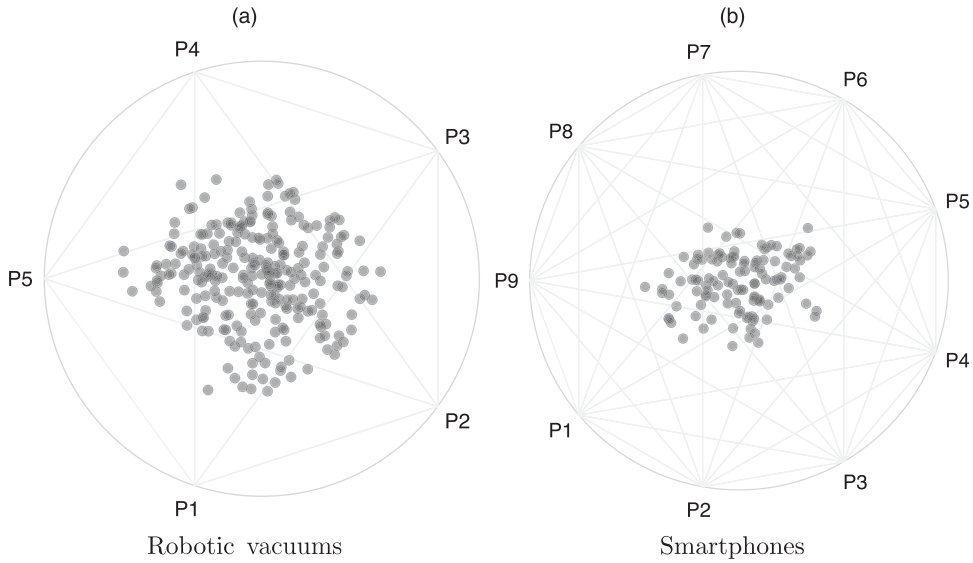
Interestingly, the VS model provides the worst predictive fit across all models for the robotic vacuum data. A tentative conclusion from this is that including interaction effects in the upper level of a hierarchical choice model does not substitute for a mixed membership approach. The reason for this result is that the regression model with (or without) variable selection considers the role of a focal covariate conditionally (i.e., while keeping all covariates fixed). In comparison, the membership vector model defines a joint probability space across covariates, and changes in the G imply changes in all covariates simultaneously.

5.2. Results from the Membership Vector Model

In the following, we present results from our proposed model. We start by looking at estimates of G. As shown in our analysis of boundary conditions in Section 3, the proposed membership vector model performs better relative to alternative models when membership in extreme profiles is truly "mixed," which implies that, a posteriori, $\alpha \gg 1$. Figure 9 displays the distribution of estimated membership vectors for our two data sets suggesting that, indeed, $\alpha > 1$.

Figure 9 reveals that for both data sets, all respondents exhibit mixed membership across all profiles, as evidenced by positions toward the center of the simplex. No respondents are positioned in the corners of the K space. Given our analysis of boundary conditions in Section 3, this result provides a partial explanation as to why the membership vector model apparently better captures preference heterogeneity:

Figure 9. Distributions of Estimated Membership Vectors



Notes. Displayed are posterior means of the G from both data sets. The simplex is defined by K extreme profiles, projected onto two dimensions. The posterior means of G clearly indicate mixed membership across respondents for both data sets.

respondents in both data sets cannot be described adequately by one of the profiles only. This also explains why the constrained membership model does not perform very well out-of-sample (Table 8).

In the following, we investigate how the proposed integrated choice and GoM model improves our ability to explain preference heterogeneity. For space considerations, we illustrate using model output from the robotic vacuum data. For details on model output for the smartphone data, the reader is referred to Online Appendix D. All results presented from the membership vector model are using ILR-transformed G (Online Appendix A) to facilitate standard interpretation of the matrix of regression coefficients Γ from this model. Recall that in the model with untransformed or "raw" G, the matrix Γ can be meaningfully applied only to full sets of g_n.

We start by considering the transposed posterior means of Γ from the binary covariate model. Table 9 displays the complete Γ' matrix. The attribute levels are on the left, and each column in the matrix is associated with the intercept or one of the statements from Table 3. The marginal posteriors with means in bold have 95% credible intervals that do not contain zero. This matrix should inform a marketer concerning the drivers of preference for promotion and targeting strategies. However, making sense of the significant values and how they might co-occur is cumbersome because coefficients are not clustered or grouped for interpretation. The parsimonious representation of the data that the profiles provide becomes increasingly important as the number of covariates included increases.

For example, we can use Table 9 to infer that respondents who are concerned about germs and dirt (i.e., Statement 5, "I worry about germs and dirt on my floor and carpet") prefer any brand of robotic vacuum relative to the outside good while not being concerned about getting the highest level of performance. We might expect this is because they are cleaning frequently (e.g., Statement 10, "I spend over two hours per week cleaning") and having a robotic vacuum is simply one part of a larger cleaning solution, but this model cannot illuminate such co-occurrence. Besides issues with interpretability, as the number of covariates increases (e.g., for the smartphone data set) the size of this coefficient matrix becomes too demanding.

In comparison, the proposed membership vector model produces a much more parsimonious coefficient matrix Γ and also provides, through $\{\lambda_j\}_{j=1}^{J}$, insights into how different items co-occur by identifying differentiated respondent profiles. Table 10 details the profiles as described by the estimates of the $\{\lambda_j\}_{j=1}^{18}$ array. Recall that because $L_j = 2$ for all J, the profiles can be presented in terms of the $\{\lambda_j(1)\}_{j=1}^{18}$

Table 9. Robotic Vacuum Binary Covariate Model Γ' Estimates

Attribute levels	Int.	S1	S2	S3	S4	S5	S6	S7	S8	S9	S10	S11	S12	S13	S14	S15	S16	S17	S18
Neato	1.52	0.03	0.22	−0.05	−0.31	**2.19**	−1.36	−0.78	**−2.18**	−0.74	−0.29	**3.75**	−1.03	1.65	−0.83	−0.66	0.56	−1.25	−0.17
iRobot	**2.76**	0.26	0.79	−0.31	−0.50	**2.41**	−1.73	−1.65	**−2.53**	−0.26	−0.26	**3.27**	−0.84	1.80	−0.95	−0.24	0.52	−0.69	0.22
Samsung	**2.96**	−0.46	0.42	−0.19	−0.23	**2.27**	−1.53	−1.02	**−1.82**	−0.76	−0.39	**3.69**	−0.95	1.37	−1.38	−0.83	1.00	−0.77	−0.27
Black & Decker	2.69	−0.42	0.21	−0.53	−0.10	**1.99**	−0.86	−1.70	**−1.85**	−0.55	−0.05	**4.23**	−1.81	**2.77**	−0.60	−0.58	0.82	−1.18	−0.07
Performance: 85%	0.91	0.43	0.29	0.47	0.68	**−1.17**	0.10	**0.93**	0.68	−0.25	0.21	−0.43	**0.92**	−0.31	0.83	0.34	0.92	−0.15	0.61
Capacity: Every 2–3 uses	0.47	−0.22	0.24	0.09	0.26	−0.02	0.21	0.07	0.03	−0.36	−0.15	−0.07	0.02	0.19	−0.10	0.14	0.43	−0.19	−0.14
Smart navigation	0.25	0.47	−0.32	0.30	0.14	0.21	0.40	0.18	0.06	0.21	0.11	−0.34	−0.21	−0.14	−0.14	−0.22	**−0.74**	0.31	−0.38
App programming	−0.50	0.14	0.48	−0.07	−0.12	0.57	−0.07	−0.10	−0.12	0.63	0.22	**−1.15**	0.05	0.34	−0.29	0.16	−0.38	−0.04	−0.21
Virtual borders	**1.01**	−0.22	−0.29	0.23	0.23	−0.10	0.07	−0.40	0.12	0.02	−0.03	−0.18	−0.08	**−1.18**	0.30	0.21	**−0.63**	0.04	0.44
$399	−1.10	−0.22	0.07	−0.36	−0.40	0.43	0.82	**−1.11**	0.73	**−1.43**	0.03	1.90	**−1.16**	−0.71	−0.07	0.65	−0.22	−0.28	−0.41
$499	**−3.04**	−0.87	0.37	0.13	−1.29	1.22	1.14	**−1.97**	1.02	**−2.54**	−0.30	3.22	**−2.93**	−1.48	0.07	**1.57**	−0.87	0.33	−0.60
$599	**−4.83**	−0.96	0.81	0.61	−1.41	0.77	0.76	−1.75	0.90	**−2.69**	−0.29	**4.28**	**−3.41**	−2.56	−0.71	**2.04**	−1.64	0.16	−0.43

Note. The marginal posteriors with means in bold have 95% credible intervals that do not contain zero; Int., Intercept; S, Statement.

Table 10. Robotic Vacuum Membership Vector Model $\{\lambda_j(1)\}_{j=1}^{18}$ Estimates

No.	Statements	$\lambda_{j,1}(1)$	$\lambda_{j,2}(1)$	$\lambda_{j,3}(1)$	$\lambda_{j,4}(1)$	$\lambda_{j,5}(1)$
1.	I enjoy coming home to a clean house.	0.96	0.94	0.94	0.75	0.52
2.	I don't feel relaxed when I know my home isn't clean.	0.87	0.88	0.83	0.11	0.12
3.	I worry about pet hair and dander in the home.	0.87	0.77	0.45	0.16	0.06
4.	I have trouble keeping the floor beneath my furniture clean.	0.85	0.47	0.74	0.15	0.08
5.	I worry about germs and dirt on my floor and carpet.	0.88	0.74	0.83	0.07	0.08
6.	I get anxious about having guests when my home is dirty.	0.89	0.9	0.89	0.19	0.09
7.	I don't like going to someone's home that is dirty.	0.85	0.86	0.91	0.12	0.05
8.	I don't like touching dirty things.	0.47	0.89	0.76	0.09	0.06
9.	I don't spend much time cleaning.	0.08	0.06	0.16	0.53	0.14
10.	I spend over two hours per week cleaning.	0.89	0.78	0.66	0.2	0.16
11.	I have a cleaning person who cleans for me.	0.04	0.07	0.15	0.04	0.06
12.	Robotic vacuums are too expensive.	0.79	0.3	0.68	0.89	0.2
13.	Robotic vacuums are too complicated to program, set up, and operate.	0.16	0.06	0.31	0.06	0.07
14.	Robotic vacuums often need to be "rescued" because they get stuck.	0.55	0.48	0.4	0.29	0.09
15.	Robotic vacuums need to have their trash containers changed too often.	0.09	0.59	0.26	0.07	0.15
16.	Robotic vacuums don't do a good enough job cleaning the floor and carpet.	0.09	0.1	0.59	0.1	0.12
17.	Robotic vacuums don't spend enough time on the really dirty spots on the floor.	0.08	0.11	0.35	0.16	0.23
18.	Robotic vacuums scare household pets.	0.32	0.24	0.2	0.17	0.08

matrix because $\{\lambda_j(0)\}_{j=1}^{18}$ is simply its complement. Because respondents were qualified by owning or being interested in a robotic vacuum, it is not surprising that nearly every profile as described in Table 10 has Statement 1, "I enjoy coming home to a clean house," occurring with high probability. Profile 1 is differentiated from the other models by Statement 10, "I spend over two hours per week cleaning," occurring with high probability. We name this profile "Constantly Cleaning."

Profile 2 is differentiated by Statement 8, "I don't like touching dirty things," along with a number of statements expressing anxiety about cleanliness with high probability. We name this profile "Anxious About Cleanliness." Profile 3, like profile 2, has high probability for a number of statements expressing anxiety about cleanliness and is differentiated by Statement 7, "I don't like going to someone's home that is dirty," and Statement 5, "I worry about germs and dirt on my floor and carpet," occurring with high probability. We name this profile "Anxious About Germs." Unlike the previously profiles, profile 4 is differentiated by Statement 9, "I don't spend much time cleaning." We simply name this profile "Little Cleaning." Finally, profile 5 is differentiated by functional concerns like Statement 17, "Robotic vacuums don't spend enough time on the really dirty spots on the floor." We name this profile "Specific Cleaning Concerns." The profile names for the robotic vacuum data are summarized in Table 11. We find similarly distinct extreme profiles with interpretable usage features for the smartphone data (Online Appendix D).

Table 12 displays the transposed matrix of estimated coefficients Γ that maps variability in the membership vectors as regressors to variability in the part-worths. Again, the marginal posteriors with means in bold have 95% credible intervals that do not contain zero. Recall that these profiles are archetypal or extreme where each individual is a partial member in each profile as defined by the weights of their membership vector g_n. To avoid the compositional nature of the original membership vectors, we imposed the ILR-transformation in Equation (13) that resulted in $g_n^{\dagger}$ for all N respondents. This transformation allows us to interpret Table 12 in the standard way and gives us a baseline that equals the preference structure of a respondent belonging equally to all latent profiles. Note that Table 12 contains a column for each of $K = 5$ profiles, which is the result of using all profiles as contrast (Online Appendix A).

As with Table 9, the matrix in Table 12 should inform a marketer concerning the drivers of preference for promotion and targeting strategies. However, using the proposed model, we are able to explain preferences in terms of the extreme profiles. For example, profile 1, Constantly Cleaning, includes Statements 5, "I worry about germs and dirt on my floor and carpet," and 10, "I spend over two hours per week cleaning," with high probability. With this profile, we can answer what was only suggested from Table 9, that an individual

Table 11. Robotic Vacuum Profile Names

No.	Robotic vacuums
1.	Constantly Cleaning
2.	Anxious About Cleanliness
3.	Anxious About Germs
4.	Little Cleaning
5.	Specific Cleaning Concerns

who is aligned with this profile prefers a high-capacity robotic vacuum relative to the outside good. In other words, because they are cleaning often, they want a robotic vacuum with high capacity in order to effectively assist in their cleaning efforts.

We can better inform targeting and promotion strategies using the proposed model by tailoring promotional messages using the groups of significant coefficients in each membership profile and targeting those respondents with high membership probabilities in a given profile. Thus, we can use the coefficient matrix as a road map for targeting by matching what respondents prefer with a far more parsimonious and richly detailed explanation of what is driving those preferences. For example, on average, consumers prefer a robotic vacuum with high cleaning performance and are very price sensitive relative to the cheapest price of \$299. For consumers who are in profile 3, Anxious About Germs, we know that price promotions should be especially effective because they have demand for robotic vacuums but are very price sensitive. The dimension reduction provided by employing an integrated choice and GoM model makes this plausible with the 12×6 Γ' matrix in Table 12 compared with a similar task using the 12×19 Γ' matrix in Table 9 from the binary covariate model.

Identifying respondent profiles is akin to a segmentation analysis. The blocks of significant attribute-level coefficients in Table 12 are reminiscent of such solutions. Unlike finite mixture models, where a respondent is assigned to a single category, mixed membership models like the GoM allow for the more realistic description of each respondent being a partial member of each profile, with weights determined heterogeneously.

6. Discussion

In this paper, we show that an integrated choice and covariate model does more to explain consumer preferences than the discrete covariates on their own. To be effective in explaining preference heterogeneity, the covariate model needs to provide individual-level summaries of the covariates, each reduced dimension to be characterized by different observations, and enable inference. An integrated choice and mixed membership model meets these criteria and outperforms the standard model and competing covariate and predictive models when respondents are a genuine mixture across profiles. In other words, mixed membership models fill the space between intercept-only and finite mixture models.

Choice modeling remains an essential fixture of marketing research. The mixed membership component of our integrated choice model uncovers both patterns across survey responses as well as how individual consumers relate to these patterns. Insomuch as these patterns or profiles relate to extremes in preference heterogeneity, and individual consumers' partial memberships in these profiles are sufficiently extreme, the integrated hierarchical Bayesian choice and GoM model (i.e., the membership vector model) uncovers distinct points of shrinkage for the model of heterogeneity. This aligns with arguments that identifying extreme responses is important for designing and promoting successful new products (Allenby and Ginter 1995). For example, extreme response behavior can be used to more efficiently target prospects with a high probability of adopting an innovation.

Conceptualizing consumer heterogeneity as a continuous distribution of preferences has been shown to aid in the identification of extreme responses (Allenby and Rossi 1998, Allenby et al. 1998). The GoM model

Table 12. Robotic Vacuum Membership Vector Model ILR-Transformed Γ' Estimates

Attribute levels	Baseline	P1	P2	P3	P4	P5
Neato	−0.22	**1.89**	**−3.05**	**1.35**	**−3.74**	**3.04**
iRobot	**1.21**	**0.88**	**−2.95**	**1.42**	**−3.49**	**3.39**
Samsung	**0.81**	**1.16**	**−2.92**	**2.10**	**−3.61**	**3.24**
Black & Decker	0.46	**1.60**	**−3.20**	**1.96**	**−3.70**	**3.31**
Cleaning Performance: 85%	**3.68**	0.15	**1.05**	**−0.71**	**0.94**	**−2.02**
Capacity: Every 2–3 uses	**0.78**	**0.26**	0.05	0.00	**−0.26**	−0.11
Smart navigation	**0.92**	**0.38**	**0.34**	−0.11	0.07	−0.15
App programming	0.01	**0.24**	**0.20**	−0.18	−0.21	0.03
Virtual borders	**0.69**	0.00	**0.51**	**−0.86**	**0.45**	0.04
\$399	**−2.09**	**−1.35**	**1.90**	**−0.89**	**−1.16**	**0.94**
\$499	**−5.60**	**−2.05**	**4.11**	**−1.46**	**−2.38**	**2.01**
\$599	**−8.56**	**−2.74**	**5.14**	**−1.87**	**−3.53**	**2.64**

Note. The marginal posteriors with means in bold have 95% credible intervals that do not contain zero.

represents discrete response behavior as a continuous proximity to a limited number of extreme profiles. Given that marketers often search for a limited number of product offerings for reasons of efficiency or resource limitations, a concept of heterogeneity that expresses differences among consumers in the space of a small number of extreme response profiles is appealing.

The empirical applications utilize typical survey and conjoint data in both an emerging market and an established category to demonstrate the use of the proposed model. However, with a growing number of potential covariates, including unstructured collections of discrete data, we see this approach as an important step to utilizing such data to improve choice modeling. LDA, as another mixed membership model, provides an obvious extension. Text data results in the same kind of sparse matrix as the multinomial data used in the GoM model, with LDA proceeding with words instead of items or statements and a single document for each individual. The dimension reduction using text is even more dramatic when starting with potentially thousands of unique words in the count matrix. However, the amount of data needed to run LDA with words composing the collection of discrete data is significant because of the large number of words in any given vocabulary. Without enough data, there are a variety of developments in topic modeling that are ripe for application within marketing, including using Dirichlet process priors (Ferguson 1973, Antoniak 1974) as a kind of distribution of heterogeneity over topic proportions. We leave the practical problems of using text in the place of traditional survey questions as an extension to this research.

Another extension relates to estimating the concentration parameter α and the optimal size of K. In working to properly account for extremes in the distribution of heterogeneity, we have seen that generating the predictive distribution of heterogeneity is sensitive to the Dirichlet hyperprior. Whereas we have minimized the influence of α in generating that predictive distribution, one might consider how to inform an additional model layer so that α can be estimated instead. Additionally, although there is not a consensus as to which measure of model fit provides the gold standard for determining the size of K, there are a number of extant methods for navigating across possible model dimensions that could be employed to include K as a parameter in the model (Green 1995, Green et al. 2015). The technical details of how to incorporate such methods into the proposed model are left for future research.

More generally, we see the use of supervised mixed membership models as an effective approach to mapping categorical responses to observed or latent outcomes of interest. This mapping is what introduces parsimony in the coefficient matrix relating the upper-level regressors and part-worths. This is also related to using mixed membership models as a model-based approach to classifying consumers that yields a more realistic description of the individual as being a mixture of various extreme consumer profiles in a way that allows us to get into the extremes of the distribution of heterogeneity. This paper serves as a step toward fulfilling a broader need to provide more complete descriptions and explanations of consumer preference heterogeneity.

References

Airoldi EM, Blei DM, Erosheva EA, Fienberg SE (2014) *Handbook of Mixed Membership Models and Their Applications*, 1st ed. (Chapman & Hall/CRC, Boca Raton, FL).

Allenby GM, Ginter JL (1995) Using extremes to design products and segment markets. *J. Marketing Res.* 32(4):392–403.

Allenby GM, Rossi PE (1998) Marketing models of consumer heterogeneity. *J. Econometrics* 89(1–2):57–78.

Allenby GM, Arora N, Ginter JL (1998) On the heterogeneity of demand. *J. Marketing Res.* 35(3):384–389.

Ansari A, Li Y, Zhang JZ (2018) Probabilistic topic model for hybrid recommender systems: A stochastic variational Bayesian approach. *Marketing Sci.* 37(6):987–1008.

Antoniak CE (1974) Mixtures of Dirichlet processes with applications to Bayesian nonparametric problems. *Ann. Statist.* 2(6): 1152–1174.

Archak N, Ghose A, Ipeirotis PG (2011) Deriving the pricing power of product features by mining consumer reviews. *Management Sci.* 57(8):1485–1509.

Blei DM, McAuliffe JD (2007) Supervised topic models. Platt JC, Koller D, Singer Y, Roweis ST, eds. *Proc. 20th Internat. Conf. Neural Inform. Processing Systems* (Curran Associates, Red Hook, NY), 121–128.

Blei DM, Ng AY, Jordan MI (2003) Latent Dirichlet allocation. *J. Machine Learn. Res.* 3:993–1022.

Büschken J, Allenby GH (2016) Sentence-based text analysis for customer reviews. *Marketing Sci.* 35(6):953–975.

Chandukala SR, Edwards YD, Allenby GM (2011) Identifying unmet demand. *Marketing Sci.* 30(1):61–73.

Clive J, Woodbury MA, Siegler IC (1983) Fuzzy and crisp set-theoretic-based classification of health and disease. *J. Medical Systems* 7(4):317–332.

Eddelbuettel D, Francois R, Allaire JJ, Ushey K, Kou Q, Russell N, Bates D, Chambers J (2018) Rcpp: Seamless R and C++ integration. Accessed August 28, 2019, https://CRAN.R-project.org/package=Rcpp.

Erosheva EA, Fienberg SE, Joutard C (2007) Describing disability through individual-level mixture models for multivariate binary data. *Ann. Appl. Stat.* 1(2):346–384.

Eugster MJA, Leisch F, Seth S (2014) archetypes: Archetypal analysis. Accessed August 28, 2019, https://CRAN.R-project.org/package=archetypes.

Ferguson TS (1973) A Bayesian analysis of some nonparametric problems. *Ann. Statist.* 1(2):209–230.

Frühwirth-Schnatter S (2006) *Finite Mixture and Markov Switching Models* (Springer-Verlag, New York).

Galyardt A (2014) Interpreting mixed membership. *Handbook of Mixed Membership Models and Their Applications* (Chapman & Hall/CRC, Boca Raton, FL), 39–65.

George EI, McCulloch RE (1993) Variable selection via Gibbs sampling. *J. Amer. Statist. Assoc.* 88(423):881–889.

Green PJ (1995) Reversible jump Markov chain Monte Carlo computation and Bayesian model determination. *Biometrika* 82(4): 711–732.

Green PJ, Łatuszyński K, Pereyra M, Robert CP (2015) Bayesian computation: A summary of the current state, and samples backward and forwards. *Statist. Comput*. 25(4):835–862.

Gross JH, Manrique-Vallier D (2014) A mixed-membership approach to the assessment of political ideology from survey responses. *Handbook of Mixed Membership Models and Their Applications* (Chapman & Hall/CRC, Boca Raton, FL), 119–139.

Horsky D, Misra S, Nelson P (2006) Observed and unobserved preference heterogeneity in brand-choice models. *Marketing Sci.* 25(4):322–335.

Jacobs BJD, Donkers B, Fok D (2016) Model-based purchase predictions for large assortments. *Marketing Sci.* 35(3):389–404.

Joutard C, Airoldi EM, Fienberg SE, Love TM (2007) Discovery of latent patterns with hierarchical Bayesian mixed-membership models and the issue of model choice. Masseglia F, Teisseire M, eds. *Data Mining Patterns: New Methods and Applications* (IGI Global, Hershey, PA), 240–275.

Kamakura WA, Russell GJ (1989) A probabilistic choice model for market segmentation and elasticity structure. *J. Marketing Res.* 26(4):379–390.

Lee SY (2007) *Structural Equation Modeling: A Bayesian Approach* (John Wiley & Sons, Hoboken, NJ).

Lee TY, Bradlow ET (2011) Automated marketing research using online customer reviews. *J. Marketing Res.* 48(5):881–894.

Lenk PJ, DeSarbo WS, Green PE, Young MR (1996) Hierarchical Bayes conjoint analysis: Recovery of partworth heterogeneity from reduced experimental designs. *Marketing Sci.* 15(2):173–191.

Manton KG, Woodbury MA, Tolley HD (1994) *Statistical Application Using Fuzzy Sets* (Wiley, New York).

Marini MM, Li X, Fan PL (1996) Characterizing latent structure: Factor analytic and grade of membership models. *Sociol. Methodology* 26:133–164.

Netzer O, Feldman R, Goldenberg J, Fresko M (2012) Mine your own business: Market-structure surveillance through text mining. *Marketing Sci.* 31(3):521–543.

Newton MA, Raftery AE (1994) Approximate Bayesian inference with the weighted likelihood bootstrap. *J. Royal Statist. Soc. B* 56(1):3–48.

Rossi PE (2018) bayesm: Bayesian inference for marketing/microeconometrics. Accessed August 28, 2019, https://CRAN.R-project.org/package=bayesm.

Rossi PE, Allenby GM (2003) Bayesian statistics and marketing. *Marketing Sci.* 22(3):304–328.

Rossi PE, Allenby GM, McCulloch RE (2005) *Bayesian Statistics and Marketing* (John Wiley & Sons, Hoboken, NJ).

Rossi PE, Gilula Z, Allenby GM (2001) Overcoming scale usage heterogeneity: A Bayesian hierarchical approach. *J. Amer. Statist. Assoc.* 96(453):20–31.

Rossi PE, McCulloch RE, Allenby GM (1996) The value of purchase history data in target marketing. *Marketing Sci.* 15(4):321–340.

Spiegelhalter DJ, Best NG, Carlin BP, Van Der Linde A (2002) Bayesian measures of model complexity and fit. *J. Royal Statist. Soc. B* 64(4):583–639.

Stewart DW (1981) The application and misapplication of factor analysis in marketing research. *J. Marketing Res.* 18(1):51–62.

Tirunillai S, Tellis GJ (2014) Mining marketing meaning from online chatter: Strategic brand analysis of big data using latent Dirichlet allocation. *J. Marketing Res.* 51(4):463–479.

Toubia O, Netzer O (2017) Idea generation, creativity, and prototypicality. *Marketing Sci.* 36(1):1–20.

Vehtari A, Gabry J, Yao Y, Gelman A (2018) loo: Efficient leave-one-out cross-validation and WAIC for Bayesian models. Accessed August 28, 2019, https://CRAN.R-project.org/package=loo.

Watanabe S (2010) Asymptotic equivalence of Bayes cross validation and widely applicable information criterion in singular learning theory. *J. Machine Learn. Res.* 11:3571–3594.

Woodbury MA, Clive J, Garson A Jr (1978) Mathematical typology: A grade of membership technique for obtaining disease definition. *Comput. Biomedical Res.* 11(3):277–298.