Management Science

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

Attribute-Level Heterogeneity

Peter Ebbes, John C. Liechty, Rajdeep Grewal

MANAGEMENT SCIENCE
Vol. 61, No. 4, April 2015, pp. 885–897
ISSN 0025-1909 (print) | ISSN 1526-5501 (online)

http://dx.doi.org/10.1287/mnsc.2014.1898

Attribute-Level Heterogeneity

Peter Ebbes
Department of Marketing, HEC Paris, 78351 Jouy-en-Josas, France, ebbes@hec.fr
John C. Liechty, Rajdeep Grewal
Department of Marketing, Smeal College of Business, Pennsylvania State University,
University Park, Pennsylvania 16802 {jcl12@psu.edu, grewalr@kenan-flagler.unc.edu}

Modeling consumer heterogeneity helps practitioners understand market structures and devise effective marketing strategies. In this research we study finite mixture specifications for modeling consumer heterogeneity where each regression coefficient has its own finite mixture—that is, an attribute finite mixture model. An important challenge of such an approach to modeling heterogeneity lies in its estimation. A proposed Bayesian estimation approach, based on recent advances in reversible-jump Markov chain Monte Carlo methods, can estimate parameters for the attribute-based finite mixture model, assuming that the number of components for each finite mixture is a discrete random variable. An attribute specification has several advantages over traditional, vector-based, finite mixture specifications; specifically, the attribute mixture model offers a more appropriate aggregation of information than does the vector specification facilitating estimation. In an extensive simulation study and an empirical application, we show that the attribute model can recover complex heterogeneity structures, making it dominant over traditional (vector) finite mixture regression models and a strong contender compared to mixture-of-normals models for modeling heterogeneity.

Keywords: heterogeneity; mixture models; hierarchical Bayes; conjoint analysis; reversible-jump MCMC; segmentation
History: Received October 5, 2010; accepted November 26, 2013, by Pradeep Chintagunta, marketing. Published online in *Articles in Advance* April 18, 2014.

1. Introduction

Since Smith (1956) first introduced the notion of market segmentation, marketing academics have devoted considerable attention to developing models to uncover heterogeneity in customers' preferences. To examine such heterogeneity, researchers often use information about consumers' demographic, economic, and psychographic variables, along with situation-specific variables such as usage occasions, to identify support points or components that could help reveal managerially relevant market segments. Finite mixture distributions can support this purpose by offering a statistical, model-based approach to segmentation (e.g., Wedel and Kamakura 2000).

Yet finite mixture distributions often appear in situations in which the identified support points or components might lack a segment-based interpretation, such as when they simply provide a flexible way to model unknown distributional shapes for unobserved heterogeneity.[1] For example, Chintagunta et al. (1991) investigate heterogeneity in brand preferences using a logit demand model and model brand choice decisions with household- and brand-specific dummies (i.e., other marketing mix effects are the same across households). For *each* brand, they use a separate finite mixture specification for the unknown heterogeneity distribution of brand preferences. Their results favor a semiparametric specification in terms of overall fit and holdout sample predictions. The estimated distributions for brand preferences differ considerably for the four brands they consider. Although each brand distribution is fit with three support points, the shapes vary from symmetric to monotonic skewed distributions. Therefore, it seems inappropriate to impose a common probability distribution for all brand preferences (Chintagunta et al. 1991).

We extend the idea of specifying a separate finite mixture distribution for each brand intercept to consider

[1] We do not intend to argue whether finite mixture components can be identified as segments in real markets; we agree with Wedel and Kamakura (2000) that this question is an empirical one. Previous studies have proposed six criteria (identifiability, substantiality, accessibility, stability, responsiveness, and actionability) to determine the appropriate basis for strategic market segmentation (e.g., Day and Wensley 1983, Wedel and Kamakura 2000). For expository convenience, we assume an implicit understanding that if mixture components support strategic segmentation, the proposals have been evaluated carefully for managerial relevance, according to the nature of competition and resources available to the decision maker. For further discussions, see also Titterington et al. (1992) and McLachlan and Peel (2000).

general finite mixture (FM) regression models that identify heterogeneity on the basis of responses to *all* marketing mix variables (e.g., Gupta and Chintagunta 1994, Kamakura and Russell 1989, Wedel and Kamakura 2000). To address the resulting estimation challenges, we adopt a Bayesian approach. From a conceptual standpoint, we note that in contemporary FM regression approaches, the optimal solution for K support points/components implies that all customer/situation variables have K components. For example, if we use information about price and loyalty to identify K components, existing FM procedures would arrive at K vector components, each with unique price sensitivities and loyalty levels. This assumption of an equal number of support points can limit the ability of FM regression models to identify complex heterogeneity structures in scenarios with different numbers of components and customer/situation variables (i.e., there are K_p price components and K_q quality components, and $K_p \neq K_q$). Furthermore, vector FM regression models, or vector models, typically can identify only around half a dozen vector components (Allenby and Rossi 1999). By relaxing the assumption of an equal number of support points across all covariates, we propose a model that identifies a different FM and heterogeneity structure for responses to each (level of an) attribute or covariate. This attribute model attains greater statistical power and recovers complex market structures in more cases because it uses data more efficiently than a traditional vector model does.

As Chintagunta et al. (1991) note, the difficulty of such an approach to modeling heterogeneity lies in its estimation. Because the model allows for varying numbers of components per attribute, the possible total number of vector components across all attributes may grow quite large. Furthermore, we do not know a priori how many support points each attribute level has. The semiparametric likelihood approach to inference, as proposed by Chintagunta et al. (1991), is not applicable in practice to an attribute model in which the attributes/covariates all have variable numbers of support points. Therefore, we develop a Bayesian estimation approach based on recent advances in reversible-jump Markov chain Monte Carlo (MCMC) methods (Brooks et al. 2003, Dellaportas and Papageorgiou 2006, Green 1995) to estimate the parameters for the attribute-based FM model. Our proposed approach solves the computational problem by placing a prior on the space of the attribute models, assuming that the numbers of components across attribute levels are discrete random variables.

We illustrate the attribute-level heterogeneity model using both simulated data and a real data set, yielding several important insights. First, the proposed attribute model strongly dominates the traditional vector regression model; it is capable of recovering much more complex heterogeneity structures than the vector model can. Second, across the conditions in our main experiment, with both continuous and discrete heterogeneity structures, the traditional vector model performs worst among all considered benchmark approaches for recovering true values of individual-level heterogeneity coefficients and making holdout predictions. The root mean squared errors (RMSEs) of the vector model are approximately 100%–150% greater on average than those of the attribute model. Third, the attribute model is fairly robust even in situations in which true heterogeneity comes from a mixture-of-normals distribution (Allenby et al. 1998, Lenk and DeSarbo 2000). A mixture-of-normals model performs best in terms of recovering the true individual-level regression coefficients in this condition (i.e., the data-generation process is a mixture of normals), but the attribute model's average RMSE is only 28% higher, whereas the average RMSE of the traditional vector model is 79% higher. The attribute model thus effectively competes with the mixture-of-normals model in the presence of continuous heterogeneity. In contrast, when true heterogeneity comes from the attribute model, the RMSE values for the mixture-of-normals model and vector model are 100% and 375% higher, respectively, than that of the attribute model. Fourth, our empirical application illustrates that the attribute model recovers a more complex heterogeneity structure, suggesting the existence of components (segments) that cannot be identified by a traditional vector model. The attribute model also offers better predictions than does the vector model in a holdout sample task and predicts similar to a mixture-of-normals model.

Our study suggests that if the researcher is agnostic about the type of heterogeneity (discrete or continuous), a simple mixture-of-normals model with one or two components may be sufficient for predictions. However, discrete FM distributions are often used as a model-based approach to segmentation (Wedel and Kamakura 2000) or as a semiparametric approach to model heterogeneity (Chintagunta et al. 1991), for which continuous-form heterogeneity models may be less useful. Our study clearly demonstrates the advantage of the attribute FM specification over the traditional FM specification for such applications.

We organize the remainder of this paper as follows: We begin by describing the attribute model in detail and contrast it with the traditional vector model. Next we discuss the computational challenges related to inferences about the attribute model and propose MCMC-based solutions. We then demonstrate differences across the various modeling approaches in practice using a simulation study and an analysis of conjoint data. We conclude with a general discussion and some potential areas for further research.

2. Attribute-Level Heterogeneity

We develop the attribute model as an alternative to the FM regression model. We illustrate the approach in a typical conjoint setting in line with our empirical application, with the standard assumption that respondents have random utility, as a function of product-specific attributes with different levels:

$$y_{it} = \beta_i' X_{it} + \varepsilon_{it}, \tag{1}$$

where y_{it} is the ith person's utility for the tth product, X_{it} is a vector of attributes (or levels of attributes in the case of discrete attributes) that describe this product, and β_i captures the ith person's partworth utilities. For our purposes, ε_{it} is independent and identically distributed and follows a normal distribution with a mean of 0 and variance of σ^2. The attribute model as described in this paper is not limited to conjoint settings and can be applied to any situation where the dependent variable y_{it} is specified as a linear function of a set of explanatory variables X_{it} so that the β_i's are the regression coefficients. In conjoint applications, the explanatory variables or covariates are often called *attributes* and the regression coefficients are often called *partworth utilities*.

The vector-based (or finite mixture) model assumes a priori that the vector of individual partworth utilities is equal to one of a set of competing mixture components (e.g., vector components) such that

$$\beta_i = \sum_{k=1}^{K} \tilde{\beta}_k I\{z_i = k\}, \tag{2}$$

where $I\{\}$ is an indicator function, z_i is a multinomial random variable that assigns each person to a mixture component, and $\tilde{\beta}_k$ is the kth vector component in the finite mixture. The vector model specification is completed by assuming a vague conjugate prior (Dirichlet prior) for the multinomial probability, where $\lambda_k = P(z_i = k)$. We set the hyperparameters for the Dirichlet priors such that a priori, each person has an equal probability of being assigned to each mixture component. For the remaining model parameters, we assume vague conjugate priors (see the Web appendix, available at http://ssrn.com/abstract=2370423, for details).

In contrast, the attribute-based model assumes a priori that each attribute partworth utility has a different set of mixture components such that

$$\beta_i(j) = \sum_{k_j=1}^{K_j} \tilde{\beta}_{k_j}(j) I\{z_{ij} = k_j\}, \tag{3}$$

where $\beta_i(j)$ is the jth partworth utility for the ith person, z_{ij} is a multinomial random variable that assigns the jth partworth for the ith person to a mixture component, and $\tilde{\beta}_{k_j}(j)$ is the k_jth mixture component for the jth attribute. Competing conjugate priors structures (Dirichlet type) can be assumed for the collection of random variables, $z_i = (z_{i1}, \ldots, z_{iK})$, which complete the prior for the attribute-based partworth vector for each individual. The marginal, independent prior assumes that the distribution of z_i is a product of prior distributions for each element of z_i; i.e.,

$$\begin{aligned} &P(z_{i1} = k_1, \ldots, z_{ip} = k_p) \\ &\quad = P(z_{i1} = k_1) \cdots P(z_{ip} = k_p) = \lambda_{1k_1} \cdots \lambda_{pk_p}, \end{aligned} \tag{4}$$

where p is the number of attributes or the size of the partworth vector β_i, and each marginal vector of probabilities λ_j is assumed to have a vague Dirichlet distribution a priori. Alternatively, the full prior does not assume independence; i.e.,

$$P(z_{i1} = k_1, \ldots, z_{ip} = k_p) = \lambda_{k_1, \ldots, k_p}, \tag{5}$$

where the vector λ (which reflects the probability of all possible combinations of the attribute-level vector that can be created by combinations of the indexes $k_1 = 1, \ldots, K_1; \ldots; k_p = 1, \ldots, K_p$) is assumed to have a vague Dirichlet distribution. We set the hyperparameters for the Dirichlet priors such that a priori, each person has an equal probability of being assigned to each mixture component. For the remaining model parameters, we also assume vague, conjugate priors (see the Web appendix for details).

Although there are important trade-offs between computational complexity and the richness of the phenomena that can be modeled using these two prior assumptions, they allow for a model that can recover the marginal posterior distribution of the partworth heterogeneity. The latter full prior specification offers a substantial computational burden because the size of the vector of probabilities λ is equal to the total number of possible combinations, $K_1 \times \cdots \times K_p$, which can become computationally prohibitive to update in an MCMC algorithm with increasing numbers of attributes and levels for each attribute. The computational problem associated with having to update the probability vector λ does not arise with the use of the marginal prior assumption in Equation (4); therefore, we used the marginal prior specification in model calibration.[2] At points, though, our discussion of the model is based on the full prior in Equation (5). In addition, in our simulation study, we test the robustness of the attribute model under the marginal independent prior specification to the full prior specification.

[2] Regardless of the prior specification for λ, there remains a substantial computational challenge associated with the model choice or the selection of the number of mixture components per attribute, as we note in the next section.

Figure 1 Illustration of Three Price Sensitivities, $\beta_i(1)$ with $K_1 = 3$, and Three Levels of Brand Loyalty, $\beta_i(2)$ with $K_2 = 3$, with All Component Mixtures on the Diagonal of Possible Combinations of Attributes ($z_{i1} = z_{i2}$ for All i)

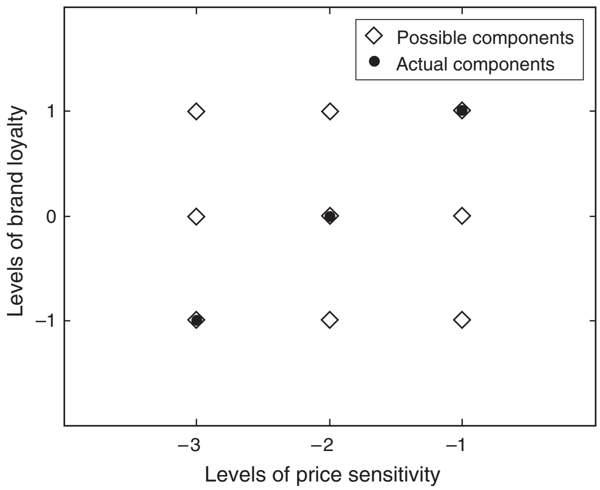
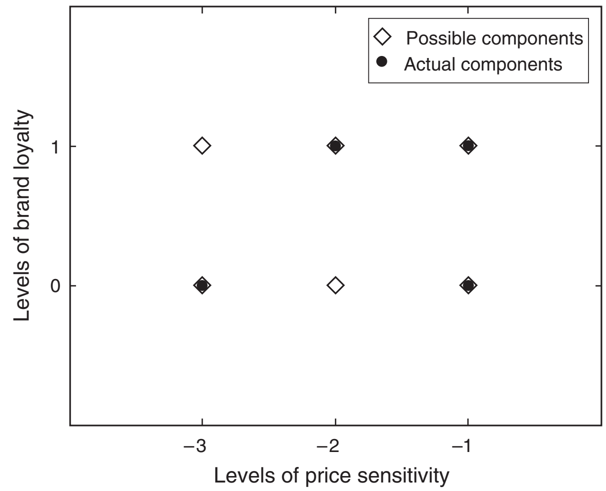


Figure 2 Illustration with Three Price Sensitivities, $\beta_i(1)$ with $K_1 = 3$, and Two Levels of Brand Loyalty, $\beta_i(2)$ with $K_2 = 2$, with Some Mixture Component Assignments Independent ($z_{i1} \neq z_{i2}$ for Some i)



To illustrate the differences between the specifications of a traditional vector FM regression model and the attribute model, we consider a simple product space described by two attributes (e.g., price and brand loyalty). The vector model implicitly assumes an equal number of components per attribute. It also assumes that knowing the component allocation of one attribute determines the component allocation of the other, and vice versa. For example, in Figure 1, three "vector" components have unique price sensitivities and loyalty levels (dark dots in diamonds), so knowing a person's price sensitivity would indicate his or her brand loyalty, and vice versa, such that $z_{i1} = z_{i2}$. The attribute model then would be redundant because it results in nine possible vector components (diamonds in Figure 1). In the redundant case, the vector model is more parsimonious than the attribute model.

However, in a more general case, a person's price sensitivity does not necessarily reveal his or her brand loyalty level, or vice versa, such as when the number of mixture components for price and loyalty is not equal. Figure 2 depicts an example with two brand loyalty components and three price sensitivity components (i.e., $K_p \neq K_q$). If a person has a price sensitivity of −1, according to the components indicated by dark dots in Figure 2, that person could have a brand loyalty of either 0 or 1. For a person with price sensitivity of −3, though, brand loyalty is 0. Thus, $z_{i1} \neq z_{i2}$ for some observations. We note that $z_{i1} \neq z_{i2}$ may arise even if K_p and K_q are equal—for instance, if we had a fourth component with a price sensitivity of −1 and brand loyalty sensitivity of 0 in Figure 1.

For this second scenario, the vector and attribute models differ in two important ways. First, the number of components with regard to representing the heterogeneity of the vector model typically increases because the number of components for price and loyalty is assumed to be equal. Figure 2 therefore contains four vector components (diamonds with dark dots) to represent just two loyalty and three price sensitivity components. The attribute model correctly takes three components for price and two components for loyalty instead of four for each. Second, the amount of information available to calibrate the parameters in the two models differs. As we illustrate in Figure 2, the number of vector components increases to four; assuming that we know how to assign respondents to components, the number of respondents used to estimate the price sensitivity of −1 or the brand loyalties of 0 and 1 will be fewer for the vector model than for the attribute model. Among those with the lowest price sensitivity (−1), the vector model splits respondents into low (0) and high (1) brand loyalty values, then estimates two price sensitivities (−1) for each brand loyalty component.[3] In contrast, such a disaggregation of information for estimating the lowest level of price sensitivity does not happen in the attribute model, which uses respondents from both low (0) and high (1) brand loyalty components to calibrate the lowest price sensitivity (−1). That is, the attribute model groups respondents by attribute, which increases the amount

[3] Empirically, the two price sensitivity estimates will be close (assuming the data are rich enough). In our experience, the estimated coefficients for one or more attributes/covariates are often similar for at least two or more components in an FM vector regression model, suggesting a more parsimonious underlying heterogeneity structure. This complexity enhances the challenge associated with identifying the underlying number of vector components in practice (e.g., Allenby and Rossi 1999).

of information available to calibrate each attribute component.[4] With these advantages, the attribute model can identify a richer heterogeneity structure than the traditional vector model can, as we demonstrate in the empirical application and simulation studies.

3. Model Estimation Challenges

Although the attribute model has advantages over the vector model, it also presents new challenges in terms of model choices and inferences (see also Chintagunta et al. 1991). The attribute model allows for a different number of mixture components for each attribute, so the possible number of mixture components can increase substantially. For example, if a conjoint study contains eight attributes and each attribute coefficient has two mixture components, as many as 2^8 $(=256)$ possible vector mixture components could exist, far more than the six components suggested as a reasonable threshold by Allenby and Rossi (1999). In real empirical settings, practitioners also do not know a priori the number of components per attribute. Even if we assume that the number is relatively small (e.g., three per attribute), it requires exploring substantial combinations of number of components (e.g., for eight attributes, 3^8 $(=6{,}561)$ possible combinations).

From an inference and model choice perspective, searching across potential combinations of a number of components, each of which is a possible model, offers a substantial computational challenge, particularly for large models. The brute-force approach would be to run an MCMC analysis for all possible models up to some fixed number of components per attribute and then calculate a model choice heuristic (e.g., Bayesian information criterion, Akaike information criterion, Bayes factor, log-marginal probability) for each model (Gamerman and Lopes 2006). A more elegant approach places a prior on the model space of all possible combinations of the number of components by making each number of components K_j a positive, discrete random variable and then uses the reversible-jump MCMC approach for inferences (Green 1995). By allowing the number of components per partworth to be random, the MCMC algorithm naturally moves around the model space, increasing or decreasing the number of components per attribute. Traditionally, MCMC algorithms that move around a variable dimension model space are known as reversible-jump MCMC algorithms (Green 1995). Richardson and Green (1997, 1998), Zhang et al. (2004), and Dellaportas and Papageorgiou (2006) have extended this basic framework to standard vector FM models; in addition, Brooks et al. (2003) generalize methods for proposing efficient algorithms. Following their general framework, we construct an appropriate reversible-jump MCMC algorithm for the attribute FM model.

A full description of the algorithm and priors appears in the Web appendix. In brief, conditional on a fixed number of components per attribute, the MCMC algorithm is straightforward and requires only the use of Gibbs sampling. The full conditional densities for the component assignment, random variables z_{ij} are multinomial, the full conditional densities for the multinomial parameters λ_{ij} are Dirichlet, the full conditional densities for each attribute-level component $\tilde{\beta}_{k_j}(j)$ are normal, and the full conditional density for the variance component σ^2 is inverse gamma.

The reversible-jump portion of the algorithm moves around the model space by increasing or decreasing the number of mixture components by 1. The algorithm either creates a new component by splitting an existing one or destroys an existing component by combining a pair of them. It might be reasonable to assume that these split and combine moves are sufficient to move around the model space, which is true in theory. However, Green (1995) finds that the mixing properties of the reversible-jump algorithm improve with the inclusion of birth and death moves. A birth move proposes a new component, with a new regression coefficient value but without any individuals assigned to it. Because there are no individuals assigned, the proposed birth component gets accepted or rejected on the basis of the prior distributions. After a birth, observations can be assigned to the new component by updating the assignment random variable z_{ij}, which tends to occur only if the proposed regression coefficient's value results in a relatively high likelihood value for the individual. The reverse of a birth is a death, which may occur only if there is an empty component with no individuals assigned to it.

4. Monte Carlo Experiment

We investigate the proposed attribute model, employing the corresponding reversible-jump MCMC algorithm, in a simulation study. We are particularly interested in the performance of the proposed approach relative to the traditional FM regression model. We start with a description of the experimental design and the factors included in the study.

4.1. Simulation Study Design

We experimentally manipulate six factors:

- F0: Data-generating process (two levels: $1 =$ attribute model or $2 =$ mixture-of-normals model).
- F1: Number of respondents (two levels: 75 or 150).
- F2: Number of questions per respondent (two levels: 10 or 20).
- F3: Number of regressors (two levels: 6 or 12).

[4] For those who exhibit the highest price sensitivity (-3), both models use the same number of respondents to estimate the price sensitivity partworth.

- F4: Structure of heterogeneity (two levels; explained subsequently).
- F5: Number of components (two levels: low and high; explained subsequently).

The first factor (F0) specifies the main data-generating process, the next two factors (F1 and F2) define the amount of information, and the last three factors (F3–F5) relate to aspects of model complexity. When the attribute model is the truth ($F0 = 1$), we generate data according to the model presented in Equations (1) and (3). Depending on the levels of F3–F5, we may obtain complex heterogeneity structures and thus can investigate whether the aggregation of information in the attribute model improves our ability to uncover the true complexity compared with a traditional vector model. When the mixture-of-normals model is the truth ($F0 = 2$), the individual-level regression coefficients are generated from a mixture-of-normals distribution—specifically, $\beta_i = \sum_{k=1}^{K} (\tilde{\beta}_k + \xi_k) I\{z_i = k\}$, where $\xi_k \sim \mathrm{N}(0, \Lambda_k)$ and Λ_k summarizes the heterogeneity covariance for the kth component. Here, $K = 1$ (low) or 3 (high), as defined by F5. The diagonal elements (within-component heterogeneity) of Λ_k are defined by F4 and equal to 0.5 or 2. When the within-component variance is low (0.5), the mixture-of-normals model approximates a vector model structure (which has within-component variance of 0). When the variances are high (2), there is a lot of within-component heterogeneity, so we can investigate the robustness of the attribute model to vector structures and the (substantial) continuous-form heterogeneity structure. The off-diagonal elements in Λ_k are not experimentally manipulated but randomly generated such that, on average, the correlations equal 0.5. Because the proposed attribute model assumes no correlations between the attribute classifications a priori, this experimental condition offers an opportunity to investigate the robustness of that assumption. Hence, the factor F4 defines the within-component variability when the mixture-of-normals model is the data-generating model ($F0 = 2$). Similarly, F4 defines the level of dependencies in the component classifications described by $z_i = (z_{i1}, \ldots, z_{iK})$ when the attribute model represents the data-generating process ($F0 = 1$). When $F4 = 1$, the classifications are generated using the full prior in Equation (5). That is, we specify nonlinear relationships between the attribute coefficients using U and L shapes. When $F5 = 1$ (low; two components per attribute), we use L shapes, and when $F5 = 2$ (high; four components per attribute), we use U shapes for each of the three ($F3 = 1$) or six ($F3 = 2$) pairs of attributes.[5] Alternatively, if $F4 = 2$, we generate classifications independently across attributes as in Equation (4). We note that $F4 = 1$ may be viewed as a simpler heterogeneity structure than $F4 = 2$ because the heterogeneity is now concentrated on fewer (joint) components (e.g., U-shaped dark dots for $F4 = 1$ versus all dark dots for $F4 = 2$ in Figure 1), so the likelihood formed from Equation (1) may be more informative for $F4 = 1$ than for $F4 = 2$. In all experimental conditions, the attribute model is calibrated using the prior specification in Equation (4).

The remaining input parameters are as follows: Across all conditions, the variance σ^2 is 1, the component locations are two points apart starting in the origin, the component sizes are uniform, and the attributes/covariates are drawn from a standard normal distribution. For the U- and L-shaped conditions, the *filled* components are of equal size.

We conduct our simulation experiment with a full-factorial design featuring $2^6 = 64$ conditions. For computational ease (for each condition, we estimated 11 Bayesian models), we use one replication per experimental condition, similar to Andrews et al. (2002).

4.2. Benchmark Models

There are two main purposes of this simulation experiment. First, we evaluate the proposed attribute model and reversible-jump MCMC algorithm as well as its robustness to various assumptions. Second, we investigate how well the traditional vector FM regression model can recover highly complex heterogeneity structures, built from simple attribute models or mixture-of-normals models. We have developed a reversible-jump approach for the vector model to consider a competing inference approach for the vector model, which essentially results in a model-averaging approach to inference. The traditional approach to inference for the vector model (e.g., Kamakura and Russell 1989, Wedel and Kamakura 2000) uses either a standard EM algorithm (likelihood optimization) or a standard MCMC algorithm (Bayesian approach) to calibrate different vector models with various components, $K = 1, 2, 3, \ldots,$ and then uses model choice heuristics to select the best one. The results we report for the vector model with reversible jump thus should be no worse than a traditional approach to inference for the vector model.

For comparison, we also compare the proposed approach with a mixture-of-normals approach to heterogeneity (Allenby et al. 1998, Lenk and DeSarbo 2000). We use up to four components,[6] including the popular shrinkage or hierarchical Bayes model with one component (e.g., Allenby and Ginter 1995, Allenby and Rossi 1999). For each choice of $K = 1, 2, 3, 4$, we estimate a mixture-of-normals model with a full variance covariance matrix Λ_k and one with no off-diagonal

[5] We thank the associate editor for this suggestion. Graphically, in Figure 1, the dark dots in the diamonds depict an L or U shape along the edges of the figure.

[6] We considered more than four components but found no major improvements in recovery and prediction. For our simulation study, a maximum of four components appears to be sufficient.

elements. For expositional clarity, we only report on the mixture-of-normals model with the highest posterior probability (Gamerman and Lopes 2006).[7]

4.3. Performance Criteria

In terms of parameter recovery, we report the RMSE for the posterior means of each respondent's partworth vector,[8] β_i, $i = 1, \ldots, n$, or RMSE(β). We also generate 10 holdout observations per respondent, which we compare with the (out-of-sample) predicted values obtained using the posterior mean estimate for β_i. We investigate the RMSE (across all respondents) by comparing the predicted and true values for the dependent variable, or RMSE(y_h). Other considered performance criteria, such as the average correlation (across attributes) between the posterior means and the true parameter values and the posterior mean of the variance σ^2, which offers a proxy for in-sample fit, gave similar insights and are omitted for brevity.

4.4. Simulation Results

To determine the statistical effects of the factors in our experiment on the two performance measures, we ran two analysis of variance (ANOVA) models (main effects and two-way interactions with estimation method only), with the factors as independent variables. Both ANOVA models are highly significant with R-squares of 0.74 and 0.71 for RMSE(β) and RMSE(y_h), respectively. All main effects are significant except for F1 (number of respondents). In contrast with the number of repeated observations per respondent (F2), the number of respondents does not strongly affect the performance of the estimation approaches. The estimation method explains most variability in the ANOVA models (highest partial η^2), followed by F3 (number of regressors) and F2 (number of repeated observations per respondent). The interaction of estimation method with F0 (data-generation process, or DGP) is the most important two-way interaction. Therefore, we report results separately for F0 = 1 (attribute model DGP) and F0 = 2 (mixture-of-normals model DGP) to contrast the performances of the considered models for heterogeneity.

Table 1 RMSE(β) for Various Conditions of Increasing Complexity, Using Data Generated from the Attribute Model (F0 = 1) and Mixture-of-Normals Model (F0 = 2)

	F0 = 1					F0 = 2				
Model	1	2	3	4	5	1	2	3	4	5
Attribute	0.03	0.04	0.12	0.11	0.60	0.31	0.31	0.40	0.42	1.10
Vector	0.14	0.75	0.67	1.96	2.29	0.57	0.62	0.89	0.89	1.15
Mixture-of-normals	0.26	0.27	0.33	0.36	1.15	0.26	0.27	0.32	0.36	0.86

Note. 1 = ideal case (smallest model, largest sample), 2 = more components, 3 = more attributes, 4 = more components and attributes, and 5 = worst case (largest model, smallest sample).

In Table 1 we report the performance of the estimation approaches in five situations with increasing complexity, including an "ideal" case (smallest model, largest sample), indicated by 1, and a "worst" case (largest model, smallest sample), indicated by 5. The patterns are similar for out-of-sample predictions. Two important conclusions emerge. First, the vector model is the worst-performing method in 9 of 10 cases. Only when F0 = 1 and in an ideal situation with the highest level of information and the smallest level of complexity is the vector model better (lower RMSE(β)) than the mixture-of-normals model, though the attribute model still achieves the lowest RMSE(β). The performance of the vector model decreases rapidly when the level of complexity increases (2 = larger number of components, 3 = larger number of attributes). At the highest level of complexity (4 and 5 in Table 1), the RMSE(β) of the vector model is approximately 18 times larger than that of the attribute model and does not appear to pick up much of the underlying heterogeneity. The attribute and mixture-of-normals models are fairly robust to increasing levels of complexity in the heterogeneity structure; only in the worst case (5 = largest model, smallest sample) do all of the models struggle. Even in this extreme case, though, it is the vector model that most dramatically fails to recover the underlying structure.

Second, when the mixture-of-normals model generates underlying heterogeneity, the attribute model does surprisingly well in terms of recovering the underlying continuous heterogeneity structure and outperforms the vector model. However, the mixture-of-normals model is the best-performing approach in all five cases and is fairly robust to increasing levels of complexity or lower levels of information. The vector model does relatively better in recovering the underlying heterogeneity for F0 = 2 than for F0 = 1, which is to be expected, because the underlying mixture-of-normals model has a vector structure. The attribute model is better able to recover the underlying continuous heterogeneity than is the vector model because it more efficiently uses and aggregates the available data. Only with the least information and the most complexity

[7] Dirichlet process priors (DPPs) have made an important contribution to modeling heterogeneity in marketing applications (e.g., Ansari and Mela 2003, Kim et al. 2004). The DPP induces a heterogeneity distribution that is a mixture of normals with a random number of components K (e.g., Van Hasselt 2011). We estimated all simulated data sets with DPPs; the results did not differ significantly from the results obtained from the mixture-of-normals approach reported in Tables 2 and 3. The detailed results for the DPP approach, and a discussion, appear in the Web appendix.

[8] An advantage of using the partworth vectors β_i (and corresponding RMSEs) is that they are not affected by the label switching possibly present in the mixture component means $\tilde{\beta}_k$ in the attribute, vector, or mixture-of-normals models.

Table 2 Simulation Study Results for F0 = 1, Including Average RMSE(β) and RMSE(y_h) by Estimation Method and Experimental Condition

	Recovery betas			Holdout sample		
Factor	Attribute	Vector	Mixture of normals	Attribute	Vector	Mixture of normals
F1						
Number of respondents						
75	0.234	1.127	0.476	1.381	3.719	1.866
150	0.225	1.053	0.456	1.334	3.503	1.799
F2						
Number of questions						
10	0.376	1.242	0.644	1.642	4.013	2.311
20	0.083	0.937	0.288	1.073	3.209	1.354
F3						
Number of regressors						
6	0.133	0.759	0.347	1.090	2.172	1.343
12	0.326	1.420	0.586	1.625	5.051	2.323
F4						
Structure						
Simple	0.178	1.062	0.456	1.228	3.579	1.819
Complex	0.281	1.118	0.476	1.486	3.644	1.846*
F5						
Number of components						
Low	0.142	0.661	0.380	1.125	2.352	1.571
High	0.318	1.518	0.552	1.590	4.870	2.094*
Overall	0.230	1.090	0.466	1.357	3.611	1.833

*The result is *not* significantly different ($\alpha = 5\%$) from the attribute model (data-generating model).

(case 5) are the attribute model and vector model similar in their performance, and both are outperformed by the mixture-of-normals model. For F0 = 1, the average RMSE(β) (five cases in Table 1) of the attribute and mixture-of-normals models is 0.18 and 0.47, respectively, and they are more dissimilar than for F0 = 2, with averages of 0.51 and 0.41, respectively. That is, the attribute model performs relatively well when the underlying heterogeneity structure comes from a mixture-of-normals model, which demonstrates the flexibility of a semiparametric specification at the attribute level and allows for the more efficient use of the available data.

We report our main results by experimental condition in Tables 2 (F0 = 1) and 3 (F0 = 2). An asterisk indicates that the result is not significantly different from that in the data-generating model. All experimental effects are generally in the anticipated direction for all models. The sample size (F1) has only a minor effect on the performance metrics for all methods, but the results improve strongly when the number of questions increases (F2). All models also perform better when the structure of the heterogeneity (F4) is less complex. In particular, the attribute model is fairly robust to situations in which the prior classifications are dependent (F4 = 1) because the difference in performance for F4 = 1 and F4 = 2 is not statistically significant for the attribute model. Hence, the influence of the likelihood dominates that of the priors with regard to the posterior, as we noted previously.

When the attribute model is the true model (see Table 2), its performance is most strongly affected by the number of questions (F2); it performs well in the presence of high levels of information. The vector model is most affected by F5 (number of components) and F3 (number of regressors). In particular, as the number of components and/or regressors increases, the ability of the vector model to recover the underlying heterogeneity structure deteriorates. The mixture-of-normals model is much less affected by the manipulated factors and performs better than the vector model in all conditions, even though the underlying heterogeneity structure is discrete (F0 = 1), which is an underlying assumption of the vector model. Overall, the average RMSE(β) for the vector and mixture-of-normals models is 374% and 103% larger, respectively, than that for the attribute model. However, the differences are much less pronounced for out-of-sample predictions, where the average RMSE improvements of the attribute model over the vector model and mixture-of-normals model are 144% and 35%, respectively. In two conditions in Table 2, indicated by asterisks, the holdout sample performance of the mixture-of-normals model does not differ significantly from that of the attribute model.

When the mixture-of-normals model is the truth (F0 = 2; see Table 3), the underlying heterogeneity structure is continuous and has a vector structure. As we might expect, the mixture-of-normals model generally dominates the attribute and vector models, and the patterns of the main effects are similar to

Table 3 **Simulation Study Results for F0 = 2, Including Average RMSE(β) and RMSE(y_h) by Estimation Method and Experimental Condition**

	Recovery betas			Holdout sample		
Factor	Attribute	Vector	Mixture of normals	Attribute	Vector	Mixture of normals
F1						
Number of respondents						
75	0.506*	0.708	0.394	1.892*	2.400	1.611
150	0.479	0.671	0.378	1.795*	2.290	1.537
F2						
Number of questions						
10	0.639	0.746	0.490	2.223*	2.496	1.824
20	0.346	0.633	0.282	1.465	2.194	1.325
F3						
Number of regressors						
6	0.395	0.607	0.332	1.402	1.797	1.296
12	0.590	0.772	0.440	2.285	2.893	1.852
F4						
Structure						
Simple	0.430	0.508	0.331	1.668	1.838	1.432
Complex	0.555*	0.871	0.441	2.020*	2.852	1.716
F5						
Number of components						
Low	0.457*	0.680	0.376	1.738*	2.302	1.545
High	0.528	0.699	0.396	1.950*	2.388	1.603
Overall	0.493	0.689	0.386	1.844	2.345	1.574

*The result is *not* significantly different ($\alpha = 5\%$) from the mixture-of-normals model (data-generating model).

the results in Table 2—with two notable observations. First, the attribute model outperforms the vector model in all cases for both recovery of heterogeneity and holdout predictions. That is, the attribute model is more robust to continuous-form heterogeneity than is the vector model, particularly when there is a lot of information (e.g., F2 = 2), which indicates that the attribute model uses information more efficiently. When F4 = 1, the attribute and vector models perform similarly, as we should expect because of the low level of within-component heterogeneity. Second, differences in performance across the three estimation models are smaller than if the attribute model were the truth (see Table 2). The vector model performs relatively better when F0 = 2 because the true heterogeneity structure has a vector structure underneath. However, the relative difference in performance between the attribute model and mixture-of-normals model is smaller for F0 = 2 than for F0 = 1. Specifically, the average improvement in parameter recovery of the mixture-of-normals model over the attribute model is 28%, and for the holdout predictions, it is 17%. In 6 of the 10 main conditions, the attribute model does not offer significantly worse predictions than the mixture-of-normals model (asterisks in Table 3).[9]

4.5. Limitations and Conclusions of Our Monte Carlo Experiment

The main conditions in our simulation experiment do not explicitly consider a case where the heterogeneity comes from a vector model. In a separate analysis (reported in the Web appendix), we find that when the vector model is the true model, it outperforms the attribute model and the mixture-of-normals model in terms of recovery of the heterogeneity and predictions, as it should. However, the attribute model still does a good job capturing the true vector heterogeneity structure. For instance, the posterior probability of correctly identifying the number of components per attribute (i.e., $K_j = K$ for all j) is 70.3% (on average across all conditions). In the cases where the number of questions per respondents is 20 instead of 10, this average increases to 90.6%. As discussed before, component classifications in the vector model are completely

[9] The attribute model recovers the true number of components per attribute with a high probability (0.88 on average) when F0 = 1. When F0 = 2, indicating continuous-form heterogeneity, the variable dimension nature of the reversible-jump attribute model correctly identifies the continuous nature of the heterogeneity structure, and the posterior average number of components per attribute tends to be large (generally five or more in our experiment), with more posterior variability. Such findings suggest that the underlying heterogeneity is probably not discrete. The high number of components explains why the attribute model can capture underlying continuous heterogeneity when F0 = 2 and is competitive with the mixture-of-normals model in that case. In addition, it highlights a diagnostic feature of the attribute model; namely, it can signal whether underlying heterogeneity is discrete or continuous.

Table 4 Attributes Used to Construct the Web Pages in the Empirical Application

Attribute	Levels		
Weather forecast	One-week (general) forecast	One-day (extensive) forecast	
University news	University sports news	General university news	
Online coupon	$2	$4	
General news	United States news (six headlines)	World News (six headlines)	Mixed news (three headlines of United States news and three headlines of world news)
Business news	General business (six headlines)	Stock market (six headlines)	Mixed news (three headlines of business news and three headlines of stock news)

dependent. We found that the attribute model can detect such dependencies by investigating the strength of the relationship or consistency between groupings in the attribute model by computing a chi-square (tail probability) value and Cramér's V statistic from cross-tabulations within each sweep of the MCMC estimation. Low values of the chi-square tail probability and high values of Cramér's V statistic indicate dependencies between the component classifications. In a simulation study we find that this procedure works as anticipated. More details are available in the Web appendix.

Collectively, the simulation and robustness studies suggest that the attribute model offers significant promise for identifying complex heterogeneity structures and is robust in situations in which the heterogeneity structure originates from a different model. With heterogeneity from the mixture-of-normals model, the attribute model performs relatively better than the vector model does. From a robustness perspective, the holdout RMSE loss reveals that the attribute model can recover enough of the underlying structure from heterogeneity to perform at a level comparable with that of the mixture-of-normals model, unlike the vector model. The vector model performs poorly in situations with relatively small amounts of information (i.e., small sample sizes) and increased complexity in heterogeneity structure.[10] The mixture-of-normals model also performs well across conditions in our experiment and is more robust than is the vector model, particularly when heterogeneity originates from the attribute model.

5. Custom Web Page Study

To determine how the attribute model performs in comparison with competing models, we collected data in an experimental setting in which respondents rated a set of Web pages that represented the customized front pages of a university news site. We administered the survey in computer labs using a computer program specifically developed for this experiment. In the final design, the Web pages included five attributes: two with three levels and three with two levels (see Table 4). Each of the 359 respondents completed a self-explicated task and a dummy task (demographic survey), then rated a set of 12 Web pages. After completing a second dummy task (university trivia quiz), two-thirds of them answered a set of demographic and product use questions, whereas the remaining respondents rated the same set of 12 Web pages again and answered the same set of demographic and product use questions. We use partially balanced blocks to which respondents were randomly assigned; a fractional factorial design for each block created the different sets of Web pages. We use dummy variable coding for the attributes in Table 4, which results in seven covariates[11] and an intercept.

We estimate the attribute and vector models for the Web study data using reversible-jump MCMC algorithms, and we estimate the mixture-of-normals models with $K = 1$, 2, 3, and 4 components and with both diagonal and full variance covariance matrices for the heterogeneity distribution. The results are similar to those from the simulation study: The vector model identifies a fairly simple heterogeneity structure with only five or six components. The attribute model uncovers a much richer heterogeneity structure, as indicated by the posterior mean of the number of components per attribute in Table 5. Finally, the mixture-of-normals model identifies three or four components.[12] From Table 5, it also follows that the assumption in the vector FM model of an identical number of

[10] The vector model often underestimates the degree of heterogeneity, but we also observed cases in which it increased the number of components to determine a solution that could capture heterogeneity across all attributes. The results thus could indicate a positive posterior probability for heterogeneity for a particular attribute, though no such heterogeneity exists, leading to "phantom" components (segments) for that attribute.

[11] The dummy variables include one-day weather forecasts, general university news, a $4 coupon, world news, mixed United States and world news, stock market news, and mixed general business and stock market news.

[12] We estimated the conjoint data with a mixture-of-normals model, with a diagonal and full variance covariance matrix and each with $K = 1$, 2, 3, 4 components. Then we used the posterior probability (Gamerman and Lopes 2006) to choose among the eight mixture-of-normals models. The largest posterior probabilities were for the model with diagonal variance covariance matrices with three and four components, at 0.33 and 0.30, respectively. We also estimated the DPP model and found it to perform similarly to the mixture-of-normals models.

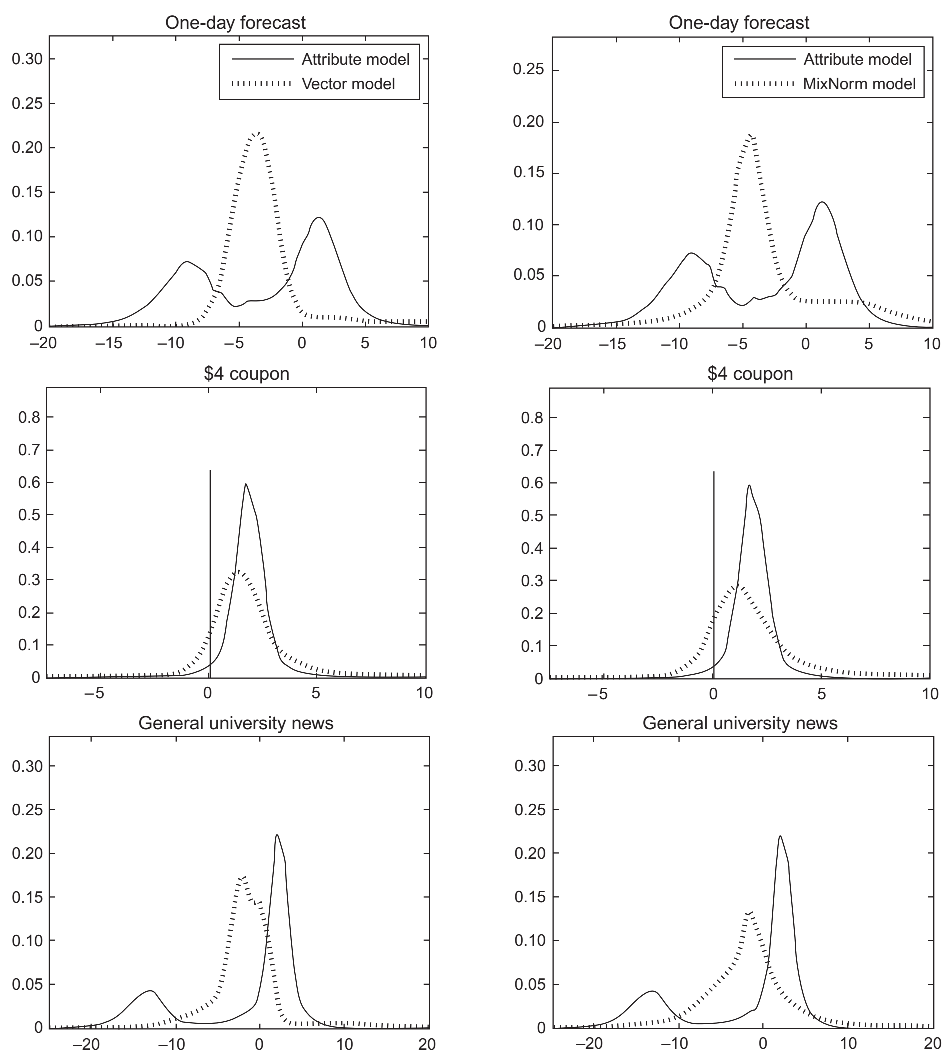
Table 5 Posterior Modes, Means, and Standard Deviations of the Number of Components per Attribute, Using the Custom Web Page Data

	Mode	Number of components posterior Mean	SD	Rounded avg. mode and mean
Intercept	2	2	0	2
One-day forecast	3	4.7	4.6	4
General university news	2	3.4	2.2	3
$4 coupon	1	1.6	1.0	1
World news	1	1.9	1.5	1
Mixed news (United States and world)	3	3.7	2.4	3
Stock market news	1	2.9	3.8	2
Mixed news (business and stock market)	1	2.9	3.6	2

support points per attribute appears unreasonable; our results highlight different heterogeneity structures across attributes.

Not only is there a stark difference in the richness of the complexity uncovered by the attribute model, but the posterior densities of the heterogeneity also differ, as we depict in Figure 3 for three attributes (the results for the other attributes are available on request). For example, in the case of the $4 coupon (which should be preferred with probability equal

Figure 3 Comparison of the Attribute Model to the Vector (Left) and Mixture-of-Normals (Right) Models: Posterior Density of Heterogeneity for Three Attributes in the Custom Web Page Study



Note. Kernel density estimates of all MCMC draws for each respondent are shown.

to 1 to the $2 coupon), the tail of the posterior distribution for the attribute model has a very small amount of probability mass for utilities that are negative, whereas the vector and mixture-of-normals models attribute an unreasonably large amount of probability mass to negative utilities. This finding provides face validity to the posterior results of the attribute model. In some cases, the competing models appear to understate the heterogeneity, such as for the one-day weather forecast or general university news. In this case, the attribute model clearly accommodates a much broader range of possible utilities, and both the vector and mixture-of-normals models miss out on the two main utility segments/components for these two attributes.

Although the attribute model indicates a much richer heterogeneity structure, it is not automatically better at describing the data; rather, the alternative structure identified by the competing models could be appropriate. We computed the posterior probabilities (Gamerman and Lopes 2006)[13] to choose among the attribute, vector, and mixture-of-normals models. The attribute model had a posterior probability of 1.0. In addition to performing better with respect to the model choice heuristic, the attribute model had the best holdout sample performance, though improvements were marginal. The improvement in mean squared error was largest (and significant) over the vector model (3.5% improvement) and smallest over the mixture-of-normals models with the largest posterior probabilities (1% for $K = 3$, not significant; 3% for $K = 4$, significant, for $\alpha = 5\%$).[14] The continuous-form heterogeneity models with one component as well the DPP approach did not perform significantly differently from the attribute model in the holdout sample. In the simulation study, the attribute and continuous-form heterogeneity models also perform similarly in terms of holdout sample performance, though relatively large in-sample improvements for recovering the (true) partworth structure could occur.

[13] In our simulation study, the approach proposed by Gamerman and Lopes (2006) correctly identified attribute versus vector heterogeneity structures in 100% of the cases. This approach was not as successful in identifying the correct number of components in the mixture-of-normals model for the data generated from this model (F0 = 2), as given by F5. If the number of components in a mixture of normals is the object of inference, we recommend a DPP approach (e.g., Ansari and Mela 2003, Kim et al. 2004).

[14] For each individual, we randomly selected five holdout ratings not used to calibrate the models. Similar to the simulation study, we used the posterior means for the partworth estimates for each individual to predict the five ratings, after which we computed the mean squared errors. We repeated this process 10 times and thus created 10 random splits of calibration and holdout samples; we then calibrated the attribute, vector, and mixture-of-normals models on the 10 data sets. This allows us to investigate variability in the holdout sample predictions.

6. Conclusions and Further Research

6.1. Conclusions

We present an alternative finite mixture model specification to identify market support points, in which each attribute-level or regression coefficient can have its own univariate finite mixture of components. This approach extends an idea put forward by Chintagunta et al. (1991), who use separate finite mixture specifications for brand dummy intercepts to capture heterogeneity in brand preferences across households. However, specifying a finite mixture for each attribute complicates estimation; we develop a reversible-jump MCMC approach (e.g., Green 1995) for that purpose. The dimension of the model space varies with the number of components, so we treat the number of components for each attribute as a discrete random variable. The joint posterior distribution of the parameters and the number of components for each attribute is the product of model probability, likelihood, and priors. An MCMC algorithm allows us to sample from the joint posterior distribution.

The attribute-level specification offers a more appropriate aggregation of information (e.g., respondents) across attributes. Using simulation studies, we demonstrate that the attribute model can recover complex heterogeneity structures with relatively small samples, such that it dominates the vector finite mixture model and competes well with the mixture-of-normals model. In our empirical application, the attribute model also dominates the vector model in terms of market insights regarding possible segments as well as predictions. However, the attribute model offers few advantages in terms of holdout sample predictions over the mixture-of-normals model in our empirical application.

Hence, if predictions are the main goal of the analysis, a simple mixture-of-normals, continuous-form heterogeneity approach may suffice. However, as Titterington et al. (1985), McLachlan and Peel (2000), or Wedel and Kamakura (2000) indicate, finite mixture distributions may be used to model heterogeneity in a cluster analysis context or can be viewed as a convenient semiparametric tool to model unknown distributional shapes of heterogeneity for which continuous form heterogeneity models may be less useful. Whether finite mixture distributions enable the identification of homogeneous market segments is an empirical question, though, that depends, among other things, on a careful evaluation of the effectiveness, efficiency, and manageability of proposals for strategic marketing segmentation (Day and Wensley 1983, Wedel and Kamakura 2000). As in the case of the vector approach, managerially relevant segments dictate the targeted marketing strategy. The main difference for targeting and positioning when using these two finite mixture approaches is that the vector model produces segments that consist of a single vector component, whereas the attribute model produces segments that are derived from a combination of one or more attribute-level components. However, the clear dominance of the attribute model over the vector model, in terms of its ability to uncover underlying heterogeneity, points out its potential usefulness for practitioners to develop market-driven segmentation strategies. Although we develop the attribute model in the context of conjoint analysis, we believe it is not limited to this particular setting, and we recommend using an attribute specification in settings when a vector FM model is considered.

6.2. Further Research

Although the attribute model shows substantial promise in practice, some of its limitations call for further research. First, we explicitly stay within the finite mixture framework and do not allow for random effects around the attribute mixture components, which would extend the (vector) mixture-of-normals distributions proposed by Allenby et al. (1998) and Lenk and DeSarbo (2000). Second, we restrict our models to a mixture of continuous, normally distributed random dependent variables (i.e., ratings data) rather than discrete dependent variables in a choice setting. This restriction can be easily lifted in theory. Our main algorithms are still valid given that many limited dependent variables may be viewed as a realization of an underlying *continuous* (latent) variable normal regression model (e.g., Rossi et al. 2006). The estimation algorithms for the considered models require intuitive and straightforward changes (e.g., considering a probit link function, where the latent values are explicitly tracked, would only require adding one additional step to the MCMC algorithms to update the latent values; details are given in the Web appendix). Given that the changes needed to adapt the MCMC algorithms for limited dependent variables are identical for the considered models, we believe that these changes were not needed to demonstrate the main differences between the vector and attribute models. Another potential extension could formalize how groupings of individuals across attributes relate, possibly through a variable selection approach (Chandukala et al. 2011). The proposed attribute model assumes a priori independent component assignments; vector model component assignments are completely dependent. A hybrid approach can be formalized by allowing for a model in which subsets of the regression coefficients group into dependent components, as a type of mini-vector model. Although these extensions would be conceptually straightforward within the proposed framework, further research is required to design efficient algorithms (e.g., parallel computing solutions) to calibrate such models.

Acknowledgments

The authors acknowledge the help of Matthew Tibbits and the feedback from Wayne DeSarbo and Peter Lenk.

Rajdeep Grewal's current affiliation is Kenan-Flagler Business School, University of North Carolina at Chapel Hill, Chapel Hill, North Carolina 27599.

References

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

Allenby GM, Rossi PE (1999) Marketing models of consumer heterogeneity. *J. Econometrics* 89:57–78.

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

Andrews RL, Ansari A, Currim IS (2002) Hierarchical Bayes versus finite mixture conjoint analysis models: A comparison of fit, prediction, and partworth recovery. *J. Marketing Res.* 39:87–98.

Ansari A, Mela CF (2003) E-customization. *J. Marketing Res.* 40(2): 131–145.

Brooks SP, Giudici P, Roberts GO (2003) Efficient construction of reversible jump Markov chain Monte Carlo proposal distributions. *J. Roy. Statist. Soc. Ser. B* 65:3–55.

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

Chintagunta P, Jain DC, Vilcassim NJ (1991) Investigating heterogeneity in brand preferences in logit models for panel data. *J. Marketing Res.* 33:417–428.

Day GS, Wensley R (1983) Marketing theory with a strategic orientation. *J. Marketing* 47:79–89.

Dellaportas P, Papageorgiou I (2006) Multivariate mixtures of normals with unknown number of components. *Statist. Comput.* 16:57–68.

Gamerman D, Lopes HF (2006) *Markov Chain Monte Carlo: Stochastic Simulation for Bayesian Inference* (Chapman & Hall/CRC, Boca Raton, FL).

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

Gupta S, Chintagunta P (1994) On using demographic variables to determine segment membership in logit mixture models. *J. Marketing Res.* 16:128–136.

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

Kim JG, Menzefricke U, Feinberg FM (2004) Assessing heterogeneity in discrete choice models using a Dirichlet process prior. *Rev. Marketing Sci.* 2:1–39.

Lenk PJ, DeSarbo WS (2000) Bayesian inference for finite mixtures of generalized linear models with random effects. *Psychometrika* 65:93–119.

McLachlan GJ, Peel D (2000) *Finite Mixture Models* (John Wiley & Sons, New York).

Richardson S, Green PJ (1997) On Bayesian analysis of mixtures with an unknown number of components. *J. Roy. Statist. Soc. Ser. B* 59:731–792.

Richardson S, Green PJ (1998) Corrigendum: On Bayesian analysis of mixtures with an unknown number of components. *J. Roy. Statist. Soc. Ser. B* 60:661.

Rossi PE, Allenby GM, McCulloch R (2006) *Bayesian Statistics and Marketing* (John Wiley & Sons, New York).

Smith WR (1956) Product differentiation and market segmentation as alternative marketing strategies. *J. Marketing* 21:3–8.

Titterington DM, Smith AFM, Makov UE (1992) *Statistical Analysis of Finite Mixture Distributions* (John Wiley & Sons, New York).

Van Hasselt M (2011) Bayesian inference in a sample selection model. *J. Econometrics* 165:221–232.

Wedel M, Kamakura WA (2000) *Market Segmentation* (Kluwer Academic Publishers, Boston).

Zhang Z, Chan KL, Wu Y, Chen C (2004) Learning a multivariate Gaussian mixture model with the reversible jump MCMC algorithm. *Statist. Comput.* 14:343–355.