On the Heterogeneity of Demand

Article *in* Journal of Marketing Research · August 1998

DOI: 10.2307/3152035

Greg Allenby
The Ohio State University

Neeraj Arora
University of Wisconsin–Madison

.... RESEARCH NOTES AND COMMUNICATIONS

GREG M. ALLENBY, NEERAJ ARORA, and JAMES L. GINTER*

Demand heterogeneity traditionally has been defined as segments of consumers that are homogeneous with regard to the benefits they seek or in their response to marketing programs (e.g., product offering, price discounts). Although it often is acknowledged that truly homogeneous segments of consumers do not exist, the approximation is assumed to be sufficiently accurate to provide a reasonable basis for the development of marketing strategy. In this article, the authors provide evidence that the homogeneous segment assumption might not be reasonable. By using a normal component mixture model that nests other, more commonly used models of heterogeneity, the authors find that the within-component heterogeneity remains substantial, even when multiple components are present. Predictive tests substantiate their finding of large within-component heterogeneity.

On the Heterogeneity of Demand

Demand heterogeneity is a critical element of marketing. It gives rise to differentiated product offerings, niche strategies, and effective target marketing efforts. During the course of the last decade, researchers have made great strides in understanding the extent of heterogeneity in many product markets. For example, latent class modeling (Kamakura and Russell 1989) now is entrenched firmly as a method in the marketing research toolbox for characterizing heterogeneous populations. This and other methods (e.g., cluster analysis) that summarize the extent of heterogeneity as a distinct number of mass points provide useful classification tools for uncovering groups of persons with similar preferences and sensitivities, often referred to in practice as *segments*.

*Greg M. Allenby is an associate professor (e-mail: allenby.1@osu.edu) and James L. Ginter is a professor (e-mail: ginter@cob.ohio.state.edu), Max M. Fisher College of Business, Ohio State University. Neeraj Arora is an assistant professor, Pamplin School of Business, Virginia Polytechnic Institute and State University (e-mail: arora@mail.vt.edu). The authors thank Peter Lenk and the anonymous *JMR* reviewers for many helpful comments.

However, because the true distribution of heterogeneity is never known a priori, there are cases in which a point mass summary might not be appropriate. For example, it might be desirable to identify and characterize the consumers in a population with preferences and sensitivities beyond some prespecified threshold. This might correspond to consumers in the tails of the distribution, those who are the least price sensitive or have the strongest preference for a particular product attribute. It is not possible to study these consumers with a finite mass approximation when the threshold value is outside the range of the estimated point masses.

In this article, we introduce a normal component mixture model of heterogeneity that nests the standard finite point approximation. We then contrast the fit of our model to a standard latent class model and find that the continuous mixture of normal distributions provides a much better fit across a variety of data sets. We then discuss substantive implications.

A FLEXIBLE MODEL FOR REPRESENTING HETEROGENEITY

We introduce the normal component mixture model in the context of a standard logit choice model in which the choice probability of alternative i for household j is

$$p_{i,j} = \frac{\exp(x_i'\beta_j)}{\sum_{\ell=1}^{L} \exp(x_\ell'\beta_j)}, \tag{1}$$

where ℓ indexes the choice alternatives, x_i is a vector of explanatory variables for alternative i, and β_j is the vector of brand preferences and sensitivities to the covariates for household j. Heterogeneity across respondents is modeled with a normal component mixture model:

$$\beta_j \sim \sum_k \phi_k \text{Normal}\left(\bar{\beta}_k, D_k\right), \tag{2}$$

where k indicates the number of components, and ϕ_k is the mass of each component. Each component is modeled with a different mean and covariance matrix (see also Lenk and DeSarbo 1997).

The normal component mixture model is capable of representing a wide variety of distributions of heterogeneity. A skewed distribution, for example, can be obtained with a two-component model with a mass point ϕ_2 small relative to ϕ_1, which results in a thicker tail near the mean of the second component. As the diagonal elements of D_k approach zero, Equation 2 converges to a finite mixture model of heterogeneity, which often is used to model homogeneous groups of consumers. The proposed model of heterogeneity is therefore very flexible because it nests the standard model and allows for the assessment of a variety of alternative forms of heterogeneity.

We estimate the model by means of the Gibbs sampler (Gelfand and Smith 1990). This estimation procedure recursively generates draws from the conditional distribution of each of the model parameters, using the realized draws of the other parameters as conditioning arguments. This recursion is executed many times, and under general conditions, the realized draws converge to draws from the true posterior distribution of model parameters. The estimation procedure yields draws from the posterior distribution of each respondent's vector of coefficients (β_j), as well as the parameters that describe the distribution of coefficients (ϕ_k, $\bar{\beta}_k$, D_k). Details of the estimation procedure are provided in the Appendix. In the subsequent analysis, we execute the recursion 10,000 times, using the first 5000 iterations as a warm-up period and the last 5000 iterations to obtain parameter estimates. (Software is available from the authors on request.)

EMPIRICAL APPLICATIONS

We compare the normal component mixture model to the latent class model with three data sets. The first data set is from a conjoint study of consumer preferences for attributes of outboard marine engines. The other two data sets are from Nielsen scanner panels of household purchases of grocery products (ketchup and tuna). These data were selected because they represent a variety of conditions for comparing the two models.

The Data

Conjoint data were obtained from a survey of boat owners in the United States. Respondents were contacted by telephone from boat registration lists maintained by state agencies. A portion of the telephone survey included a conjoint exercise designed to estimate consumer preferences for brand names and sensitivities to various outboard engine features. A list of the attributes and attribute levels considered in the analysis appears in Table 1.

We used the "trade-off" method of conjoint data collection. Each respondent indicated their preference for various attributes by choosing between two engines that were identical in every respect, except that the first engine had a low value on attribute a and a high value on attribute b, whereas the second engine had a high value on a and a low value on b. For example, the first engine was less fuel efficient but more reliable, whereas the second engine was described as more fuel efficient but less reliable. Specific attribute levels and brand names were used to collect the data but are disguised here for proprietary reasons. A total of 7500 observations from 543 respondents were available for analysis.

Table 2 provides a description of the scanner panel data sets. The scanner panel data differ from the conjoint data in that the outcomes are multinomial rather than binomial, and explanatory variables, such as price, are continuous rather

Table 1

DESCRIPTION OF THE OUTBOARD ENGINE DATA

Sample Size:
543 Respondents
7500 Observations

Attributes:
1. Brand Name
   Six brands of outboard engines were studied: A, B, C, D, E, and F.
2. Price
   x% less (more) expensive.
3. Fuel Economy
   x% more (less) fuel efficient.
4. Reliability
   Starts quickly every (x% of the) time.
5. Durability
   x% decreased (increased) risk of mechanical failure.
6. Vibration and Noise
   x% less (more) vibration and noise.
7. Acceleration
   Gets on "plane" x% faster (slower).
8. Speed
   x% faster (slower) for the same horsepower.
9. Emissions
   x% less (more) smoke.
10. Technology
   Cutting-edge (standard) technology.

Table 2

DESCRIPTION OF SCANNER PANEL DATA

Brand	*Choice Share*	*Average Price ($)*	*Proportion of Occasions* *Displayed*	*Featured*
Ketchup (735 households, 4312 observations)				
Heinz	.431	1.226	.083	.142
Hunt's	.239	1.268	.103	.056
Del Monte	.106	1.279	.054	.012
House	.224	.775	.085	.052
Tuna (209 households, 4643 observations, ≥15 observations/household)				
Chicken of the Sea (water)	.395	.712	.217	.352
Starkist (water)	.304	.769	.133	.255
House (water)	.066	.631	.119	.155
Chicken of the Sea (oil)	.126	.722	.175	.292
Starkist (oil)	.108	.759	.218	.236

than discrete. Two product categories, ketchup (in 32-ounce bottles) and tuna (in 6-ounce cans), were selected for analysis because they represent different levels of purchase frequency. Households purchase ketchup less frequently than tuna, which results in fewer observations and greater statistical error in the assignment of households to components. The tuna data have, on average, 22 records per household, whereas the ketchup data have 6. Both data sets were obtained from ACNielsen scanner panels in Springfield, Mo. The selection of households in the tuna data set was restricted to those with more than 15 purchases during a two-year period. This later data set is therefore representative of heavy purchasers for whom the statistical error of parameter estimates is smaller.

In-Sample and Predictive Fit

Parameter estimates were obtained for the normal component mixture model (Equations 1 and 2) and a latent class model that constrains D_k to be zero in Equation 2. Table 3 reports in-sample and predictive fits for these models. Two observations per respondent were reserved as holdout samples for the outboard engine data. For the ketchup data, one observation per household was reserved for predictive testing because of the short purchase histories available for each household. Two observations per household were reserved for predictive testing of the tuna data.

In-sample fit for the latent class model was measured with Schwarz's (1978) information criterion, which asymptotically approximates the marginal density of the data. The marginal density is used to construct the Bayes factor (see Allenby 1990) used in Bayesian hypothesis testing. For the hierarchical Bayes model, the marginal density was calculated using the harmonic mean of the likelihood values evaluated over the iterations of the Gibbs sampler (see Newton and Raftery 1994).

The predictive fit is based on the mean absolute deviation (MAD) between the observed choices and the associated choice probabilities, evaluated at the mean of the posterior distribution of each respondent's (household's) parameter estimates (for the mixture model, see Kamakura and Russell 1989, p. 381; for the hierarchical Bayes model, see Allenby and Ginter 1995). The MAD was selected because of its ease of interpretation; a MAD of .00 indicates a perfect set of predictions, whereas a MAD of .50 indicates a predictive performance equivalent to a coin flip in a binary setting.

In Table 3, the best log marginal density and predictive MAD for both latent class and hierarchical Bayes models are underlined. For comparison purposes, the best statistics for each model type are used. The in-sample and predictive fit statistics indicate strong support for the hierarchical Bayes model rather than the finite mixture model. The log marginal density is 45% smaller for the outboard engine data, 30% smaller for the ketchup data, and 24% smaller for the tuna data. This indicates that the extent of unobserved heterogeneity is not captured adequately by the finite mixture model (with a small number of mass points) and that it is not correct to constrain D_k to be zero in Equation 2. This result is confirmed in the predictive tests, in which the predictive MAD for the hierarchical Bayes model is smaller by 14% for the outboard engine data, 25% for the ketchup data, and 25% for the tuna data. The finite mixture model does not require estimating the $k \times d \times (d + 1)/2$ parameters associated with the k covariance matrices D_k, each of dimension $d \times d$. Therefore, many more mass points are needed for the finite mixture model to capture the extent of heterogeneity adequately. In the next section, we examine the extent of within- and across-component heterogeneity in the three data sets.

In addition to indicating model fit and appropriateness, the pattern of results directly affects managerial interpretation. In each case, the number of components associated with the latent class model was identified with a standard fit statistic, the Bayesian Information Criterion (Schwarz 1978). Yet in each case, the in-sample and predictive fits are much worse, which indicates the need for either additional components or a more flexible model, such as a normal component model. Therefore, it does not appear that procedures based on the standard latent class model provide a sound basis for identifying segments.

Parameter Estimates

In general, parameter estimates for the conjoint and scanner panel data indicate that the extent of within-component heterogeneity is much greater than the extent of across-component heterogeneity. The component means ($\bar{\beta}$) and unobserved heterogeneity estimates (diagonal elements of D) are generally greater than twice their posterior standard deviation, which indicates that the estimates are statistically different from zero. Parameter estimates are not reported here but are available on request. Instead, we provide a graphical summary of the estimates by plotting the aggregate distribution of heterogeneity. This aggregate distribution is

Table 3

IN-SAMPLE AND PREDICTIVE FIT

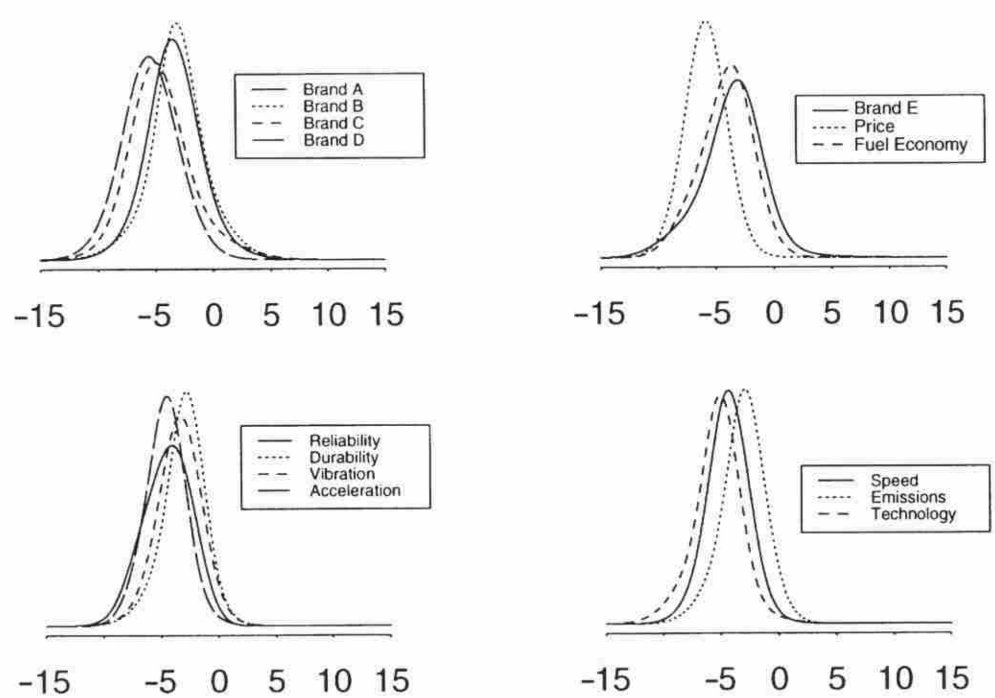
	Latent Class Model		*Normal Component Model*	
Data Set/ Components	*Log Marginal Density*[a]	*Predictive MAD*[b]	*Log Marginal Density*	*Predictive MAD*[b]
Outboard Engine				
1 Component	−4040.5	.481	−2351.8	.421
2 Component	−3842.5	.478	−2322.9	.411
3 Component	<u>−3783.3</u>	.467	−2171.7	.407
4 Component	−3784.0	<u>.454</u>	−2161.3	.404
5 Component	−3797.8	.455	<u>−2072.0</u>	<u>.396</u>
6 Component	−3949.0	.455	−2082.5	.405
Ketchup				
1 Component	−3096.1	.469	−2000.3	.324
2 Component	−2797.3	.421	−1956.4	.311
3 Component	−2726.1	.399	−1941.0	.309
4 Component	<u>−2708.0</u>	.387	<u>−1901.5</u>	<u>.305</u>
5 Component	−2720.6	<u>.381</u>	−1918.9	.309
Tuna				
1 Component	−4499.0	.467	−2588.5	.271
2 Component	−3859.5	.388	−2567.9	.270
3 Component	−3521.1	.364	−2560.3	.267
4 Component	−3451.8	.362	−2560.4	.268
5 Component	−3370.9	.345	−2568.4	.266
6 Component	−3344.6	.347	−2567.7	.264
7 Component	<u>−3306.6</u>	.332	<u>−2520.5</u>	<u>.261</u>
8 Component	−3310.7	<u>.327</u>	−2547.4	.262

[a]Approximated with the Bayes Information Criteria for the finite mixture model: Log-likelihood − parameters/2 ln(observations).

[b]Two observations holdout per respondent for outboard engine and tuna data, one observation for ketchup.

Note: MAD = mean absolute deviation.

Figure 1
DISTRIBUTION OF HETEROGENEITY—OUTBOARD ENGINES



obtained by weighting each of the component distributions by the estimated masses, in accordance with Equation 2. If the extent of across-component heterogeneity is greater than the extent of within-component heterogeneity, the aggregate distribution will exhibit distinct multimodality.

Figure 1 displays plots of the distribution of heterogeneity for the outboard engine data. None of the 14 aggregate distributions in the figure shows signs of distinct multimodality, which indicates that the extent of within-component heterogeneity is much greater than the extent of across-component heterogeneity.

Figure 2 plots estimates of the distribution of heterogeneity for the ketchup data. The plots show greater evidence of multimodality for the brand intercepts. The left tails of the distributions in the upper half of Figure 2 are composed of households that did not purchase the corresponding brand. This result is consistent with the presence of consideration sets (McAlister and Srivastava 1991; Roberts and Lattin 1991), which have been shown to be a major aspect of preference heterogeneity. It should be noted that this aspect is not present in conjoint tasks in which respondents are forced to consider alternative product offerings.

Figure 3 displays results for the tuna data. Recall that households included in the analysis were required to have more than or equal to 15 observations. Although few households (21 of 209) have complete purchase histories, in that they were observed to purchase each of the choice alternatives, the many observations allows for a more accurate assessment of the homogeneity restriction for the price, display, and feature coefficients. We detect some evidence of multimodality in the distribution for price, though the degree of within-component heterogeneity is still great. There continues to be evidence of much greater within- to among-component heterogeneity for the display and feature coefficients.

In summary, examination of the distributions of heterogeneity reveals much within-component heterogeneity in consumer preferences for product features, brand names, and sensitivity to price, displays, and feature advertising. Estimates of unobserved heterogeneity are large, relative to the extent of across-component heterogeneity. Furthermore, these estimates have much greater predictive validity than those derived from conventional latent class models that assume no within-component heterogeneity.

DISCUSSION

The results of our analysis provide evidence that the extent of heterogeneity is much greater than that measured by latent class models. The three data sets represent different marketing information domains, from stated preferences based on partworths to observed brand preferences in the marketplace. Yet, we find the distribution of heterogeneity to be continuous in each. For all parameter estimates in the three data sets, the extent of within-component heterogeneity typically is estimated to be greater than the extent of across-component heterogeneity, which results in distributions of heterogeneity for which well-defined and separated modes do not exist. In other words, across the three data sets we investigated, a discrete approximation did not appear to characterize the marketplace completely or accurately.

One possible explanation for the heterogeneity observed in our analyses is that markets are too large and too complex to be represented adequately as a few homogeneous market segments. Because of all the benefit combinations and usage scenarios, a much larger sample of segments might be re-

Figure 2
DISTRIBUTION OF HETEROGENEITY—KETCHUP

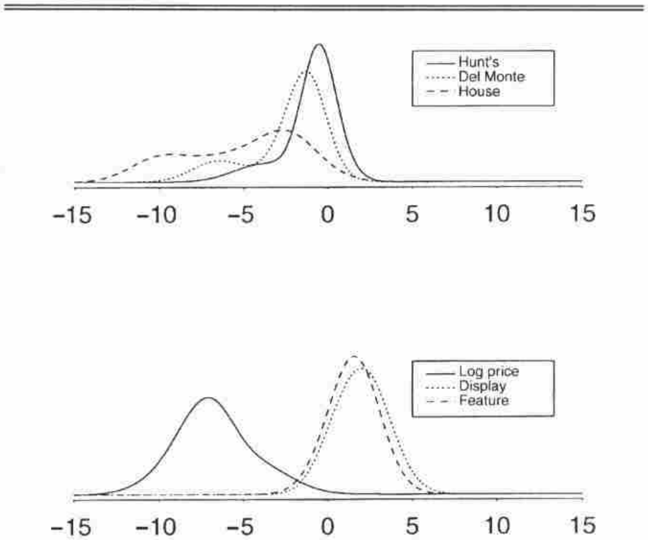
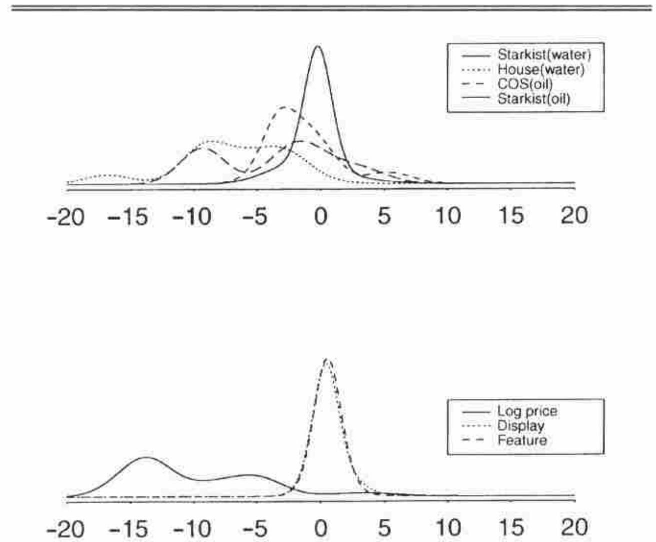


Figure 3
DISTRIBUTION OF HETEROGENEITY—TUNA



quired to represent the market. A physical analogy of this situation might be an attempt to model a mountain range (e.g., the Rockies) as a few density functions. Although the range consists of many somewhat separable mountains, it cannot be represented well as a small number of large mountains. Any such solution attempt would exhibit the physical equivalent of (within-segment) heterogeneity and would be arbitrary and incorrect. A better representation of this market would be to consider many segments, perhaps a hundred or more.

The more complete picture of the distribution of demand provided by the normal component model is important when considering decisions that are nonlinearly related to the distribution of heterogeneity (see also DeSarbo and Grisaffe 1996). We note, for example, a repositioning decision in which a change in the quality of a product is being considered. An important component of this decision is the demand for quality among those persons who are price insensitive, because potential profits exponentially increase as price sensitivity approaches zero. It is therefore important to fully characterize and use the entire distribution of price sensitivity in this problem rather than employ the summary measures provided by the latent classes. Because profits are nonlinearly related to the distribution of heterogeneity, there is no guarantee that the latent class approximation will lead to good estimates of expected profits or optimal product configurations.

Recent advances in estimation methods, such as the Gibbs sampler, now permit estimation of complicated model structures that previously were not possible to estimate (Allenby and Ginter 1995; Rossi, McCulloch, and Allenby 1996). These methods offer a rich and more complete understanding of the distribution of demand, which is often not available from a particular parametric density (e.g., normal) or a small sample of mass points. For many managerial decisions, it is worth the researcher's effort to estimate this distribution so that marketing efforts can be designed and directed toward the most profitable consumers.

APPENDIX

Estimation Algorithms

The Gibbs sampler for the model described by Equations 1 and 2 involves recursively generating random draws from the full conditional distribution of the model parameters. For a restricted model with one component distribution of heterogeneity (k = 1 in Equation 2), the recursion becomes

1. Generate $\{\beta_j, j = 1, ..., J\}$:

$$f\left(\beta_j \middle| \bar{\beta}, D\right) \propto |D|^{-1/2} \exp\left[-1/2\left(\beta_j - \bar{\beta}\right)' D^{-1}(\beta_j - \bar{\beta})\right] \times \prod_m p_{i,j,m},$$

where m indexes the m choices for respondent j. Draws were obtained using the Metropolis-Hastings algorithm with a random walk chain (see Chib and Greenberg 1995). That is, denote $\beta^{(p)}{}_j$ as the previous draw for β_j. The next draw, $\beta^{(n)}{}_j$, is then given by

$$\beta_j^{(n)} = \beta_j^{(p)} + \Delta\beta,$$

where $\Delta\beta$ is a draw from the density Normal(0, .25). The choice for parameters of this density ensures an acceptance rate of greater than 50%. The probability of accepting this new draw is given by

$$\text{Prob(acceptace)} = \min\left\{\left[\exp -\tfrac{1}{2}\left(\beta_j^{(n)} - \bar{\beta}\right)' D^{-1}\left(\beta_j^{(n)} - \bar{\beta}\right)\right] \prod_m p_{i,j,m}\left(\beta_j^{n}\right) \Big/ \left[\exp -\tfrac{1}{2}\left(\beta_j^{(p)} - \bar{\beta}\right)' D^{-1}\left(\beta_j^{(p)} - \bar{\beta}\right)\right] \prod_m p_{i,j,m}\left(\beta_j^{p}\right), 1\right\},$$

where $p_{i,j,m}(\beta_j)$ denotes the choice probability evaluated at β_j.

2. Generate $\bar{\beta}$:

$$f\left(\bar{\beta}\middle|\{\beta_j\}, D\right) = \text{Normal}\left(\sum_{j=1}^{J} \beta_j/J, D/J\right).$$

3. Generate D:

$$f\left(D\middle|\{\beta_j\}, \bar{\beta}\right) = \text{Inverted Wishart}$$

$$\left[\sum_{j=1}^{J}\left(\beta_j - \bar{\beta}\right)\left(\beta_j - \bar{\beta}\right)' + G, J + g\right],$$

where G and g are prior parameters that are set to G = 15I and g = 15. The Gibbs sampler for the model with one-component heterogeneity distribution (k = 1) proceeds by recursively executing steps 1 through 3.

The normal component mixture model can be estimated with a Gibbs sampling procedure by introducing a latent variable, s_{jk}, which is used to index the k components to which each of the j respondents belong. On each iteration of the procedure, only those respondents assigned to a component are used to arrive at the component mean ($\bar{\beta}_k$) and covariance matrix (D_k). The assignment of respondents to components is straightforward if $\{\phi_k\}$, the mass points of the components, are known.

$$\Pr\left(s_{jk} = k'\right) = \left\{\phi_{k'}(2\pi)^{-p/2}\left|D_{k'}\right|^{-1/2} \exp\left[-1/2\left(\beta_j - \bar{\beta}_{k'}\right)'\right.\right.$$

$$\left.D_{k'}^{-1}\left(\beta_j - \bar{\beta}_{k'}\right)\right]\Big/\sum\nolimits_k \phi_k(2\pi)^{-p/2}\left|D_k\right|^{-1/2}$$

$$\left.\exp\left[-1/2\left(\beta_j - \bar{\beta}_k\right)' D_k^{-1}\left(\beta_j - \bar{\beta}_k\right)\right]\right\},$$

where p is the dimension of β. That is, $\Pr(s_{jk} = k')$ is evaluated for all components, normalized to add to one, and then assigned by comparing this probability with a random draw from a uniform (0,1) distribution. This procedure is executed on each iteration of the Gibbs sampling procedure for each respondent.

To obtain ϕ_k, we assume a Dirichlet prior, and given $\{s_{jk}\}$, this leads to a Dirichlet posterior with prior parameters revised to reflect the number of respondents assigned to each component. A random draw from this posterior is obtained by first generating draws from gamma distributions and then normalizing these draws. That is, for each component (k), we generate $z_k \sim \text{gamma}(\Sigma_j s_{jk} + \alpha_k, 1)$ and then set $\phi_k = z_k/\Sigma_\ell z_\ell$, where α_k denotes the prior parameters. In our analysis, we set $\alpha_k = 10$ for all k. Given the large sample sizes present in the three empirical studies, this prior specification results in a mild influence on the posterior.

Achieving model identification is a common problem when using normal component mixture and latent class models (see Diebolt and Robert 1994; McLachlan and Basford 1988). This is because the likelihood is identical for different labeling of the components. For example, in a two-component (k = 2) mixture model, it does not matter whether $\beta' = (\beta_1', \beta_2')$ and $D = \{D_1, D_2\}$ or $\beta' = (\beta_2', \beta_1')$ and $D = \{D_2, D_1\}$, because the results are identical. We achieve statistical identification in our model by imposing an ordinal restriction on the component masses: $\phi_1 > \phi_2 > \ldots > \phi_k$ through the prior distribution. The implication of this assumption is to accept draws of $z_k \sim \text{gamma}(\Sigma_j s_{jk} + \alpha_k, 1)$ only, for which $\phi_k = z_k/\Sigma_\ell z_\ell$, k = 1, ..., K follows the assumed ordering. In addition, the use of proper prior distributions for the covariance matrices avoids the problem of an unbounded likelihood function, because the prior bounds the covariance matrix away from zero.

REFERENCES

Allenby, Greg M. (1990), "Hypothesis Testing with Scanner Data: The Advantage of Bayesian Methods," *Journal of Marketing Research*, 27 (November), 379–89.

——— and James L. Ginter (1995), "Using Extremes to Design Products and Segment Markets," *Journal of Marketing Research*, 32 (November), 392–403.

Chib, Siddartha and Edward Greenberg (1995), "Understanding the Metropolis-Hastings Algorithm," *The American Statistician*, 49 (4), 327–35.

DeSarbo, Wayne S. and D. Grisaffe (1996), "Combinatorial Optimization Approaches to Constrained Market Segmentation: An Application to Industrial Market Segmentation," working paper, School of Business, University of Michigan.

Diebolt, Jean and Christian P. Robert (1994), "Estimation of Finite Mixture Distributions Through Bayesian Sampling," *Journal of the Royal Statistical Society, Ser. B*, 56 (2), 363–75.

Gelfand, Alan E. and Adrian F. M. Smith (1990), "Sampling-Based Approaches to Calculating Marginal Densities," *Journal of the American Statistical Association*, 87 (June), 523–32.

Kamakura, Wagner A. and Gary J. Russell (1989), "A Probabilistic Choice Model for Market Segmentation and Elasticity Structure," *Journal of Marketing Research*, 26 (November), 379–90.

Lenk, Peter J. and Wayne S. DeSarbo (1997), "Bayesian Inference of Finite Mixture, Random Effects, and Generalized Linear Models," working paper, School of Business, University of Michigan.

McAlister, Leigh and Rajendra Srivastava (1991), "Incorporating Choice Dynamics in Models of Consumer Behavior," *Marketing Letters*, 2, 241–52.

McLachlan, Geoffrey J. and Kaye E. Basford (1988), *Mixture Models: Inference and Applications to Clustering*. New York: Marcell Drekker Inc.

Newton, Michael A. and Adrian E. Raftery (1994), "Approximate Bayesian Inference with the Weighted Likelihood Bootstrap," *Journal of the Royal Statistical Society (B)*, 56 (March), 3–48.

Roberts, John H. and James M. Lattin (1991), "Development and Testing of a Model of Consideration Set Composition," *Journal of Marketing Research*, 28 (November), 429–40.

Rossi, Peter E., Robert E. McCulloch, and Greg M. Allenby (1996), "The Value of Purchase History Data in Target Marketing," *Marketing Science*, 15, 321–40.

Schwarz, Gideon (1978), "Estimating the Dimension of a Model," *The Annals of Statistics*, 6, 461–64.