RICK L. ANDREWS, ANDREW AINSLIE, and IMRAN S. CURRIM*

Currently, there is an important debate about the relative merits of models with discrete and continuous representations of consumer heterogeneity. In a recent *JMR* study, Andrews, Ansari, and Currim (2002; hereafter AAC) compared metric conjoint analysis models with discrete and continuous representations of heterogeneity and found no differences between the two models with respect to parameter recovery and prediction of ratings for holdout profiles. Models with continuous representations of heterogeneity fit the data better than models with discrete representations of heterogeneity. The goal of the current study is to compare the relative performance of logit choice models with discrete versus continuous representations of heterogeneity in terms of the accuracy of household-level parameters, fit, and forecasting accuracy. To accomplish this goal, the authors conduct an extensive simulation experiment with logit models in a scanner data context, using an experimental design based on AAC and other recent simulation studies. One of the main findings is that models with continuous and discrete representations of heterogeneity recover household-level parameter estimates and predict holdout choices about equally well except when the number of purchases per household is small, in which case the models with continuous representations perform very poorly. As in the AAC study, models with continuous representations of heterogeneity fit the data better.

# An Empirical Comparison of Logit Choice Models with Discrete Versus Continuous Representations of Heterogeneity

In a recent study, Andrews, Ansari, and Currim (2002; hereafter, AAC) compared the relative effectiveness of models with discrete versus continuous representations of consumer heterogeneity in the context of metric conjoint analysis. They compared finite mixture (FM) models, which describe heterogeneity with discrete distributions, and hierarchical Bayesian (HB) estimation of models with continuous representations of heterogeneity in terms of parameter recovery, fit, and prediction. The study found that FM- and HB-estimated models were equally effective in recovering individual-level parameters and predicting ratings of holdout profiles, but HB-estimated models fit the data better than did FM models. In addition, both model specifications were shown to be quite robust to violations of underlying assumptions. For example, an HB model with a unimodal prior performed well even when true partworths came from a mixture of distributions, and FM produced very good parameter estimates at the individual level, despite its intended usage at the segment level.

The goal of the current study is to compare the relative empirical performance of logit choice models with discrete versus continuous representations of heterogeneity in experimental conditions similar to those of the AAC study. As yet, the relative empirical advantages and disadvantages of using a discrete or continuous distribution to represent consumer heterogeneity in choice model applications are unknown. Wedel and colleagues (1999, p. 223) state that "To a large extent, the issue of a continuous vs. a discrete distribution of heterogeneity is an empirical one…. We need more simulation studies … and more empirical studies to fully understand the strengths and weaknesses of several methods for estimating household-level parameters."

*Rick L. Andrews is an associate professor, Department of Business Administration, University of Delaware (e-mail: andrewsr@udel.edu). Andrew Ainslie is an assistant professor, Anderson School of Management, University of California, Los Angeles (e-mail: andrew.ainslie@anderson.ucla.edu). Imran S. Currim is Corporate Partners Research Scholar and a professor, Graduate School of Management, University of California, Irvine (e-mail: iscurrim@uci.edu). The authors contributed equally to the article. Rick Andrews acknowledges the support of a research grant from the University of Delaware.



*Journal of Marketing Research*
Vol. XXXIX (November 2002), 479–487

Given the different distributional assumptions, model forms, and model estimation procedures of metric conjoint analysis and logit choice models, it is anything but a forgone conclusion that the findings of the AAC study will hold in choice model applications. Also, because the information content of choice data is less than that of metric ratings data, the potential for accurately recovering unobserved consumer heterogeneity from the data may be lower as well. We could speculate that, given the lower information content of choice data, models with more sophisticated continuous representations of heterogeneity may not have any advantage over models with less sophisticated representations and may even perform worse if spurious heterogeneity effects or overfitting occurs.

In this study, FM logit models (e.g., Kamakura and Russell 1989), which have been applied extensively to scanner panel data, are used to recover discrete distributions of heterogeneity. We use mixed logit models estimated with HB estimation methods to recover continuous distributions of heterogeneity. Mixed logit models, also known as random coefficients logit models, have received considerable attention in marketing and econometrics (e.g., Brownstone and Train 1999; Chintagunta, Jain, and Vilcassim 1991; Elrod 1988; Erdem 1996; Gönül and Srinivasan 1993; Revelt and Train 1998). McFadden and Train (2000) establish that, under mild regularity conditions, mixed logit can approximate as closely as needed the choice probabilities of any discrete choice model derived from random utility maximization. In the past decade, HB-estimated models have emerged with much potential for representing heterogeneity in consumer preferences (e.g., Allenby and Rossi 1999; McCulloch and Rossi 1994; Rossi, McCulloch, and Allenby 1996). Collectively, the FM model and mixed logit model estimated with HB represent the state of the art for representing heterogeneity in logit choice models. Although these procedures have been applied in the marketing literature for approximately a decade, no study has compared and contrasted their parameter recovery, fit, and forecasting accuracy in experimentally controlled conditions. The study by Huber and Train (2001) compares a mixed logit model estimated with simulated maximum likelihood with one estimated with an HB procedure and finds virtually equivalent individual-level parameter estimates, so we do not report the results of mixed logit models estimated with simulated maximum likelihood estimation.[1] The study by Abramson and colleagues (2000) compares a variety of logit models in several experimentally controlled conditions, but the models estimated were not compared with their HB counterparts.

The next section describes the data generation process and the models fit to the simulated data sets. We then present the results and discuss their implications.

## *DESIGN OF THE MONTE CARLO STUDY*

### *Data*

The simulation design (choice of factors and levels) is influenced by studies conducted by Vriens, Wedel, and Wilms (1996; hereafter, VWW) and AAC. The AAC study provides details of the experimental design. Seven factors were experimentally manipulated for this study:

1. The number of mixture components: one, two, or three;
2. The separation between mixture components: small (.5) or large (1.0);[2]
3. The distribution of coefficients within components: normal or gamma;
4. The variance of coefficients within components: .05 or .25;[3]
5. The number of households: 75, 200, or 400;[4]
6. The number of purchases per household: 3, 10, or 15; and
7. The error variance: standard (1.645) or large (2 × standard, or 3.290).[5]

Factors 5 and 6 are observable to the analyst prior to model estimation, whereas the other factors are not. The number of households (Factor 5) and number of purchases per household (Factor 6) are representative of the sample sizes observed in actual empirical applications using scanner panel data, though the levels are purposely designed to be somewhat extreme at the endpoints so that model robustness may be investigated. For most of the unobservable factors, the levels were set through experimentation, again with the goal of creating manipulations that test the robustness of the various models (for details on the choice of levels, see nn. 2–5). Identifying models that are most robust to variations in data characteristics is valuable, especially when the characteristics are unobservable. We can use a robust model with more confidence that it is accurately recovering choice behavior, regardless of the characteristics of a particular data set.

Regarding the number of mixture components (Factor 1), we expect HB-estimated mixed logit models to be preferred over FM when there is only one component (at least when the distribution of coefficients within components is normal), because the FM model is designed to handle multiple components but not within-component heterogeneity. The other data sets will have two or three components, and each component will have a distribution of coefficients, resulting in multimodal distributions for coefficients. Both models are underspecified with two- or three-component data; FM does not model within-component heterogeneity, and HB-

[1]These results are available from the authors on request.

[2]Previous research has manipulated the separation between segment means by multiplying all coefficients by two (see VWW and AAC). This type of manipulation increases the separation between segments. However, in the context of multinomial logit models, this manipulation of separation has the undesirable consequence of changing the scale factor of the model, which makes conclusions about parameter recovery more difficult. Andrews and Currim (2003) find that the manipulation of segment separation by changing the average separation between segments (.5 and 1.0) is an improvement. Differences of .5 between segments are fairly difficult to detect, whereas differences of 1.0 are much easier to detect.

[3]The higher end variance is larger than that in the VWW and AAC studies (both considered variances of .05 and .10). The variance is increased because previous studies find that this manipulation is weak for some measures of performance. We made the manipulation stronger to ensure a thorough understanding of the effects of within-component variance.

[4]VWW and AAC used (100, 200) and 150 consumers, respectively. We used a larger range in this study to test the robustness of the models at smaller sample sizes and the convergence of the models at larger sample sizes.

[5]The variance of the extreme value distribution is $\pi^2\theta^2/6$, which is 1.645, assuming the typical scale factor value of $\theta = 1$. Note that multiplying the error variance by two is equivalent to dividing the parameter values by $\sqrt{2}$. This is because the scale factor is confounded with the parameter values.

estimated mixed logit does not model more than one component. It is an empirical question and a test of robustness as to how well these models will perform in the two- and three-component conditions.

Likewise, we expect FM models to perform better relative to HB when the separation between components is larger (Factor 2). When the average difference between components is smaller, the HB-estimated models may adequately handle the multimodality in preferences and responses to marketing activities, but this is less likely when the average difference between components is larger.

There is speculation that models with continuous distributions are susceptible to misspecification of the functional form of the distribution of heterogeneity (Heckman and Singer 1984; Wedel et al. 1999). Our third experimental factor produces data sets with normal or gamma distributions of heterogeneity. Whereas normal distributions produce bell-shaped curves, gamma distributions are highly skewed. Because of the shape of the gamma distribution, we expect that FM models will perform better than HB-estimated mixed logit models when there is gamma heterogeneity within components.

The variance of within-component distributions (Factor 4) is either .05 or .25. The larger the within-component variance, the better the performance of HB is expected to be relative to FM. When the variance is smaller, it is possible that the FM models will adequately recover heterogeneity, but this is an empirical question.

Factors 5 and 6 manipulate the number of households in the sample and the number of purchases per household. In the context of conjoint analysis, there is evidence that HB methods can recover heterogeneity and estimate individual-level parameters even when individual-level least squares estimators do not exist because of insufficient degrees of freedom (Lenk et al. 1996). However, we have included an extreme case in which there are three observations and seven parameters per household to determine if the HB method breaks down under extreme conditions. We might expect HB methods to perform better than alternative methods when there are fewer households in the sample (Wedel et al. 1999), but this too is an empirical question. The FM and HB models should converge to similar estimates when there are many households (e.g., 400) and/or many purchases per household (e.g., 15).

The AAC study finds that HB estimation of a conjoint analysis model with continuous heterogeneity distributions does not produce good results when the error variance (Factor 7) is large. It is not clear that this finding will hold in the context of logit models applied to scanner panel data. Although all models should perform better with smaller error variance, we will be determining whether the performance of the HB-estimated mixed logit models deteriorates as the error variance increases.

Because Factor 2 (separation of components) is not meaningful when there is only one preference component in the data, the design is not a full factorial. When there are two or three components in the data, there are $2^5 3^2 = 288$ experimental conditions. When there is one component, there are $2^3 3^2 = 72$ experimental conditions, for a total of 360 conditions. Similar to VWW and AAC, we have one data set per experimental condition, resulting in 360 data sets. With four models estimated per data set (described in the next section), the simulation produces $360 \times 4 = 1440$ observations. However, because of the interpretation of Factor 2 when there is one preference component, only $288 \times 4 = 1152$ observations can be used in the statistical analysis of variance (ANOVA) of the simulation results. The power of this design to detect small differences between model types (a small effect accounts for approximately 1% of the variance of the dependent variable) at a significance level of .05 is approximately 81% (Cohen 1988). For factors with two levels (1, 2, 3, 4, and 7), the power to detect small differences at the .05 level is approximately 92%. For factors with three levels (5 and 6), the power to detect small differences is approximately 87%.

The generation of the choice data closely follows related simulation studies, such as VWW and AAC. Essentially, the predictor variables include one continuous variable (e.g., price) and two binary variables (e.g., store feature advertisement and aisle display). The continuous variable was generated from a standard normal distribution; the binary variables were generated such that they have values of 1 approximately 10% of the time, similar to the promotion frequency observed in actual scanner panel data. We assume that there are five alternatives in the consumer's choice set.

The true parameters for each data set were randomly generated from a uniform distribution, and the ranges of the parameters are typical of those observed in actual scanner panel applications. The brand-specific constants are generated to be in the range of –1.5 to 1.5, the coefficients of the price variable are in the range of –1 to –2.5, and the coefficients of the promotion variables are in the range of 1 to 2.5. The brand-specific constants are smaller in absolute value than the other parameters so that there would be adequate choices of each alternative to ensure identification and accurate estimation of all parameters. The sizes of the components are also generated randomly for each data set, and constraints are set up such that no component is smaller than 10% of the sample and none is larger than 90% of the sample (see Cutler and Windham 1994).[6] By generating the parameters independently for each data set, we have more assurance that the simulation results generalize beyond a specific set of parameter values chosen by the researcher. Regardless of whether each consumer has 3, 10, or 15 purchases used for model estimation (Factor 6), 5 purchases per consumer are used for model validation.

### *Models*

*FM models*. Conditional on the vector of coefficients for segment s, $\beta^s$, the probability of household h's sequence of choices $y_h$ is as follows:

$$l\left(y_h \middle| \beta^s\right) = \prod_{t=1}^{T_h} \prod_{k=1}^{K} P(y_{hkt} = 1 | X_{ht}, \beta^s)^{y_{hkt}}, \tag{1}$$

where $y_{hkt}$ is the 0/1 choice indicator variable, and

[6]The relative masses of components was experimentally manipulated in the AAC study, but the factor did not affect the relative performance of HB and FM models.

$$(2)\qquad P\left(y_{hkt} = 1 \middle| X_{ht}, \beta^s\right) = \frac{\exp\left[V_{ht}^s(k)\right]}{\sum_j \exp\left[V_{ht}^s(j)\right]},$$

where $V_{ht}^s(k)$ is household h's utility for brand k at purchase occasion t if in segment s. The unconditional probability of observing household h's sequence of choices is

$$(3)\qquad l(y_h|\Theta) = \sum_s \alpha_s l\left(y_h|\beta^s\right),$$

where $\Theta$ represents all the parameters $\beta^s$, s = 1, …, S, and the mixing weights $\alpha = (\alpha_1, \alpha_2, \ldots, \alpha_s)$ are interpreted as segment sizes, such that $0 < \alpha_s < 1$ and $\Sigma_s\alpha_s = 1$. The log-likelihood for all households is

$$(4)\qquad LL = \sum_h \ln\ l(y_h|\Theta).$$

The parameters are estimated through maximization of Equation 4 through numerical optimization.

After estimating the parameters, we estimated household-level parameter vectors $\beta_h$ using a weighted average of the segment-level parameter vectors $\beta^s$, for which the weights are the posterior probabilities of segment membership. The posterior probability for household h belonging to segment s is computed as

$$(5)\qquad \pi\left(\hat{\beta}^s|y_h\right) = \frac{\hat{\alpha}_s l\left(y_h\middle|\hat{\beta}^s\right)}{\sum_{s'} \hat{\alpha}_{s'} l\left(y_h\middle|\hat{\beta}^{s'}\right)},\ s = 1, \ldots, S.$$

The estimates of household-level parameters are then

$$(6)\qquad \hat{\beta}_h = \sum_s \pi\left(\hat{\beta}^s|y_h\right)\hat{\beta}^s.$$

Household-level parameters estimated in this manner were used in the computation of parameter estimation error, fit, and forecasting accuracy, as discussed subsequently in this section.

One issue with the estimation of FM models is how to determine the number of components S to retain in the models. One approach, used by VWW, is to retain the true number of components. However, this approach may give an unfair advantage to FM because it involves the use of information that will not be available to analysts in real-world settings. However, because there is heterogeneity within each component, this approach may penalize FM because components beyond the true number might be required to account for the heterogeneity. For example, AAC used the Bayesian information criterion (BIC) to identify the number of components and found that BIC often fit more than the true number of components. A study by Andrews and Currim (2003) has shown that the validation sample log-likelihood is far superior to BIC for identifying the number of components to retain for FM logit models. In the current study, we consider the performance of two specifications of FM models: one in which the number of components is determined by BIC (FM-B) and the other in which it is determined by the validation sample log-likelihood (FM-L).

*HB-estimated mixed logit models*. Unlike classical estimation approaches, the HB approach considers the joint distribution of the household-level parameters and the common parameters. The joint (posterior) distribution of the household-level parameters can be written as follows (Allenby and Rossi 1999; Wedel et al. 1999):

$$(7)\qquad \pi\left(\beta_h, \Theta|y_h\right) \propto l\left(y_h|\beta_h\right)f\left(\beta_h|\Theta\right)f(\Theta),$$

where the three terms after the proportionality sign are the likelihood, the mixing distribution, and the prior for $\Theta$, respectively. Inferences about the preferences and sensitivities of a specific household are obtained by appropriately marginalizing this joint posterior distribution (Allenby and Rossi 1999):

$$(8)\qquad \pi\left(\beta_h|y_h\right) = \int \pi\left(\beta_h, \Theta|y_h\right)\ d\beta_{-h}\ d\Theta,$$

where –h refers to all households except h.

For the covariance matrix on the parameters, we chose an inverse Wishart prior, $W[\rho, (\rho R)^{-1}]$, where $\rho$ is the number of parameters plus two and R = diag(.1), consistent with the AAC study on which this study is based. In each case, the Monte Carlo Markov Chain sampler was run for 20,000 iterations, 5000 of which were used for burn-in. Likelihood and time-series checks demonstrated that this was more than adequate.

*Homogeneous logit models*. As a benchmark for the heterogeneous logit models, we estimate homogeneous logit models for all data sets. Comparison of the heterogeneous FM and mixed logit models estimated with HB with the homogeneous models will provide information on the levels of improvement achieved by explicitly modeling consumer heterogeneity.

### *Measures of Performance*

Similar to VWW and AAC, we use several measures to assess the performance of the models in terms of parameter recovery error, fit, and predictive accuracy. The root mean square error (RMSE) between the actual and estimated household-level parameters, RMSE($\beta$), is computed as a measure of parameter recovery error, as in VWW and AAC. The log-likelihood and marginal log-likelihood are used to assess model fit. In the case of the FM and logit models, BIC was used as an approximation to the marginal likelihood. The BIC is based on the Schwartz (1978) criterion, which in turn is an approximation of the marginal likelihood of a model. For the HB model, the marginal likelihood was computed using the reweighted importance sampling method outlined by Raftery (1996) and Newton and Raftery (1994). To make this comparable to the BIC, a statistic widely used in the marketing literature, we compared –2(marginal log-likelihood) to the BIC measure for the other models. Hereafter, we refer to both as –2MLL. Finally, we used two measures to assess prediction accuracy: the log-likelihood values from the validation samples and the average predicted choice probability for the chosen brand, also computed from the validation samples.

## *RESULTS OF THE MONTE CARLO STUDY*

The results are summarized in Tables 1, 2, and 3. Table 1 contains the overall ANOVAs for the five dependent vari-

Table 1

F-TESTS OF MAIN AND INTERACTION EFFECTS ON PERFORMANCE MEASURES

| *Source (d.f.)* | *F-Ratio RMSE(β)* | *F-Ratio LOGL(E)* | *F-Ratio –2MLL* | *F-Ratio LOGL(V)* | *F-Ratio $\overline{P}$(V)* |
|---|---|---|---|---|---|
| Model | 64.03 | 190.17 | 47.87 | 2.62 | 25.94 |
| (3) | (<.0001) | (<.0001) | (<.0001) | (.0495) | (<.0001) |
| Number of components | 223.48 | .03 | 10.00 | 5.38 | 3.42 |
| (1) | (<.0001) | (.8669) | (.0016) | (.0205) | (.0646) |
| Separation between components | 322.50 | 12.75 | .09 | .14 | 1.89 |
| (1) | (<.0001) | (.0004) | (.7628) | (.7094) | (.1693) |
| Within-components distribution | 1.61 | 6.27 | 5.27 | 7.82 | 8.06 |
| (1) | (.2046) | (.0124) | (.0218) | (.0053) | (.0046) |
| Within-components variance | 529.79 | 6.86 | 15.78 | 3.59 | 11.74 |
| (1) | (<.0001) | (.0089) | (<.0001) | (.0585) | (.0006) |
| Number of households | 82.65 | 1.66 | 17.32 | 10.03 | 6.7 |
| (2) | (<.0001) | (.1899) | (<.0001) | (<.0001) | (.0013) |
| Number of purchases/household | 191.33 | 8.06 | 15.00 | 10044.70 | 5.99 |
| (2) | (<.0001) | (.0003) | (<.0001) | (<.0001) | (.0026) |
| Error variance | 715.74 | 502.37 | 397.20 | 199.51 | 487.49 |
| (1) | (<.0001) | (<.0001) | (<.0001) | (<.0001) | (<.0001) |
| Model × components | 25.75 | 5.56 | 2.20 | .99 | 2.85 |
| (3) | (<.0001) | (.0009) | (.0869) | (.3952) | (.0365) |
| Model × separation | 34.30 | 6.30 | 2.74 | 1.01 | 3.90 |
| (3) | (<.0001) | (.0003) | (.0420) | (.3874) | (.0087) |
| Model × within-components distribution | 1.17 | .02 | .01 | .06) | .00 |
| (3) | (.3208) | (.9970) | (.9980) | (.9798 | (1.0000) |
| Model × within-components variance | 12.80 | 1.16 | .67 | .40 | .31 |
| (3) | (<.0001) | (.3226) | (.5723) | (.7566) | (.8172) |
| Model × households | 18.03 | 4.26 | .30 | 1.41 | .33 |
| (6) | (<.0001) | (.0003) | (.9353) | (.2062) | (.9233) |
| Model × purchases/household | 27.69 | 13.50 | .77 | 1.92 | .83 |
| (6) | (<.0001) | (<.0001) | (.5964) | (.0742) | (.5439) |
| Model × error variance | 10.05 | 2.41 | 1.17 | .19 | 1.65 |
| (3) | (<.0001) | (.0655) | (.3211) | (.9011) | (.1762) |
| $R^2$ | .73 | .53 | .37 | .95 | .37 |

Notes: The ANOVAs are based only on data sets with two or three preference components because the separation of components factor is not meaningful when there is only one component. n = 1152, and *p*-values are in parentheses; d.f. = degrees of freedom.

ables, Table 2 contains the means of the parameter recovery error measure RMSE(β) and the measures of fit (the estimation sample log-likelihood LOGL[E] and the marginal log-likelihood –2MLL), and Table 3 contains the means of the prediction measures (the validation sample log-likelihood LOGL[V] and the average choice probability $\overline{P}$[V]). All likelihood values and –2MLL are expressed on a per-observation basis to improve the interpretation of Factors 5 and 6. At the bottom of Tables 2 and 3, we have computed the overall means by model type and tested these means for significant differences using the ANOVA results in Table 1. The main results from the experiment are described subsequently.

One of the major findings we observe from Tables 2 and 3 is that the HB-estimated models perform very poorly in the three-purchases-per-household condition. Parameters are poorly identified at the level of the individual because there are seven parameters but only three purchases, and this appears insufficient to allow the shrinkage across households to form good estimates. Although the fit measures LOGL(E) and –2MLL are extremely good for HB models in this condition, parameter recovery RMSE(β) and forecasting accuracy LOGL(V) are both extremely poor, which is indicative of overfitting. According to these measures, HB performs worse than the homogeneous logit model. The average choice probability $\overline{P}$(V) does not indicate poor performance for HB in the three-purchase condition, but compared with LOGL(V), this measure is much less sensitive to very low predicted choice probabilities.[7] Because all the other factor-level means and the overall means for HB methods in Tables 2 and 3 are contaminated by the poor numbers for the three-purchases-per-household condition, we reserve further judgment on the performance of the HB method compared with other methods, pending further analysis. We first focus on the performance of the other methods in Tables 2 and 3.

With regard to parameter recovery (Table 2), the FM models have the best RMSE(β) means overall. These models recover household-level parameters significantly better than does the homogeneous logit model. For the RMSE(β) means in the interior of Table 2, differences of approximately .0238 are required for statistical significance (ignoring the adjustment for multiple comparisons). In general, according to the overall factor means, the models recover parameters better when there are fewer components in the data, the separation between components is smaller, the

[7]The range of household-level parameter estimates is greater for HB models, which could translate into a greater range of predicted choice probabilities for the chosen brand and possibly some values near zero. Taking the log of such a small probability results in a severe penalty for LOGL(V), but this is much less so for the average choice probability because no logs are taken.

Table 2

PARAMETER RECOVERY (RMSE(β)) AND FIT (LOGL[E] AND –2MLL) MEANS, BY MODEL TYPE AND EXPERIMENTAL CONDITION[a]

| | *RMSE(β)* | | | | | *LOGL(E)*[d] | | | | | *–2 MLL*[e] | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| *Factor* | *FM-B*[c] | *FM-L* | *HB* | *LOGIT* | *Mean* | *FM-B* | *FM-L* | *HB* | *LOGIT* | *Mean* | *FM-B* | *FM-L* | *HB* | *LOGIT* | *Mean* |
| *Number of components*[b] | | | | | | | | | | | | | | | |
| 1 | .3443 | .3473 | .4254 | .3464 | .3659 | –1.0071 | –.9893 | –.8287 | –1.0128 | –.9595 | 2.0753 | 2.0835 | 1.9439 | 2.0766 | 2.0448 |
| 2 | .4371[1] | .4355[1] | .4997[1] | .4835[1] | .4640[1] | –.9890 | –.9685 | –.8083 | –1.0269[1] | –.9482 | 2.0865 | 2.0939 | 1.9171 | 2.1049[1] | 2.0506[1] |
| 3 | .5114 | .5034 | .5399 | .6637 | .5546 | –.9710 | –.9462 | –.7900 | –1.0805 | –.9469 | 2.1374 | 2.1470 | 1.9083 | 2.2120 | 2.1012 |
| *Separation* | | | | | | | | | | | | | | | |
| .5 | .4333[1] | .4327[1] | .4872[1] | .4661[1] | .4548[1] | –1.0041 | –.9824 | –.8208 | –1.0384[1] | –.9614 | 2.1153 | 2.1230 | 1.9470 | 2.1279 | 2.0783 |
| 1.0 | .5151 | .5063 | .5524 | .6810 | .5637 | –.9559[1] | –.9322[1] | –.7775[1] | –1.0690 | –.9337[1] | 2.1086 | 2.1178 | 1.8784[1] | 2.1890 | 2.0735 |
| *Within-component distribution* | | | | | | | | | | | | | | | |
| Normal | .4753 | .4687 | .5334 | .5751 | .5131 | –.9690 | –.9483 | –.7903 | –1.0436 | –.9378[1] | 2.0933 | 2.1016 | 1.8969 | 2.1383 | 2.0575[1] |
| Gamma | .4731 | .4702 | .5062[1] | .5721 | .5054 | –.9910 | –.9663 | –.8080 | –1.0637 | –.9573 | 2.1307 | 2.1392 | 1.9285 | 2.1785 | 2.0942 |
| *Within-component variance* | | | | | | | | | | | | | | | |
| .05 | .3914[1] | .3848[1] | .4817[1] | .5002[1] | .4395[1] | –.9666 | –.9493 | –.7994 | –1.0342[1] | –.9374[1] | 2.0752[1] | 2.0813[1] | 1.9005 | 2.1195[1] | 2.0441[1] |
| .25 | .5571 | .5542 | .5579 | .6470 | .5790 | –.9935 | –.9654 | –.7989 | –1.0731 | –.9577 | 2.1487 | 2.1595 | 1.9249 | 2.1974 | 2.1076 |
| *Number of households* | | | | | | | | | | | | | | | |
| 75 | .5129 | .5124 | .6473 | .5808 | .5633 | –1.0053[2] | –.9834[2] | –.7584[1] | –1.0613 | –.9521 | 2.1817 | 2.1951 | 1.9488[2] | 2.2139[2] | 2.1349 |
| 200 | .4636[1] | .4572[1] | .4664[1] | .5787 | .4915[2] | –.9776[1,2] | –.9562[1,2] | –.8189 | –1.0593 | –.9530 | 2.1049[1] | 2.1116[1] | 1.9173[1,2] | 2.1586[1,2] | 2.0731[2] |
| 400 | .4461[1] | .4388[1] | .4457[1] | .5613 | .4730[1] | –.9571[1] | –.9324[1] | –.8202 | –1.0404 | –.9375 | 2.0493[1] | 2.0545[1] | 1.8721[1] | 2.1028[1] | 2.0197[1] |
| *Number of purchases/household* | | | | | | | | | | | | | | | |
| 3 | .5495 | .5430 | .6856 | .5891[2] | .5918 | –1.0059[2] | –.9821 | –.7053[1] | –1.0543[1,2] | –.9369[1] | 2.1927 | 2.1983 | 1.9377 | 2.2034[2] | 2.1330 |
| 10 | .4541[2] | .4492[2] | .4600[2] | .5600[1] | .4808[2] | –.9566[1] | –.9299[1] | –.8218[2] | –1.0362[1] | –.9361[1] | 2.0561[1] | 2.0665[1] | 1.8771 | 2.1066[1] | 2.0681[2] |
| 15 | .4191[1] | .4163[1] | .4138[1] | .5717[1,2] | .4552[1] | –.9775[1,2] | –.9600 | –.8703 | –1.0705[2] | –.9696 | 2.0872[1] | 2.0964[1] | 1.9233 | 2.1653[1,2] | 2.0266[1] |
| *Error variance*[f] | | | | | | | | | | | | | | | |
| Standard | .5456 | .5410 | .5913 | .6835 | .5904 | –.8870[1] | –.8616[1] | –.7089[1] | –.9844[1] | –.8605[1] | 1.9523[1] | 1.9621[1] | 1.7318[1] | 2.0199[1] | 1.9165[1] |
| High | .4028[1] | .3980[1] | .4483[1] | .4637[1] | .4282[1] | –1.0730 | –1.0530 | –.8894 | –1.1230 | –1.0346 | 2.2717 | 2.2787 | 2.0936 | 2.2970 | 2.2352 |
| Overall | .4742[1] | .4695[1] | .5198[2] | .5736[3] | .5093 | –.9800[3] | –.9573[2] | –.7992[1] | –1.0537[4] | –.9475 | 2.1120[2] | 2.1204[2,3] | 1.9127[1] | 2.1584[3] | 2.0759 |

[a]Superscripts on overall means by model type at the bottom of the table indicate significant differences (according to least significant difference rule) at the .05 level, with the superior mean having a value of "1." In the interior of the table, factor level means within model types are tested for significant differences in a similar manner. The RMSE values used in this analysis, from the ANOVAs in Table 1, are .1029 (RMSE[β]), .1318 (LOGL[E]), and .2714 (–2MLL).

[b]The means for one-component data sets were not analyzed in the ANOVAs in Table 1 and therefore are not tested for significance in this table.

[c]FM-B = finite mixture models with the number of components determined by BIC; FM-L = finite mixture models with the number of components determined by LOGL(V); HB = mixed logit models estimated by hierarchical Bayes procedure; and LOGIT = homogeneous logit models.

[d]The log-likelihood is the posterior likelihood, computed from household-level parameter estimates. All likelihoods are divided by the number of observations in the data set to facilitate interpretation.

[e]Approximated by BIC for FM and LOGIT models and divided by the number of observations to facilitate interpretation.

[f]In the high error variance condition, the true parameter estimates were divided by the square root of 2 before computation of RMSE(β).

Table 3

PREDICTION ACCURACY (LOGL[V] AND $\overline{P}$[V]) MEANS BY MODEL TYPE AND EXPERIMENTAL CONDITION[a]

| | *LOGL(V)*[d] | | | | | $\overline{P}$*(V)*[e] | | | | |
|---|---|---|---|---|---|---|---|---|---|---|
| *Factor* | *FM-B*[c] | *FM-L* | *HB* | *LOGIT* | *Mean* | *FM-B* | *FM-L* | *HB* | *LOGIT* | *Mean* |
| *Number of components*[b] | | | | | | | | | | |
| 1 | –.8591 | –.8580 | –.8851 | –.8598 | –.8655 | .4714 | .4746 | .4921 | .4700 | .4770 |
| 2 | –.8583 | –.8559 | –.8874 | –.8662[1] | –.8669[1] | .4760 | .4800 | .4978 | .4644[1] | .4795 |
| 3 | –.8750 | –.8701 | –.8935 | –.9136 | –.8880 | .4746 | .4801 | .4973 | .4386 | .4726 |
| *Separation* | | | | | | | | | | |
| .5 | –.8684 | –.8660 | –.8943 | –.8745 | –.8758 | .4691 | .4732 | .4919 | .4599[1] | .4735 |
| 1.0 | –.8649 | –.8600 | –.8866 | –.9052 | –.8792 | .4815 | .4869 | .5032 | .4430 | .4786 |
| *Within-component distribution* | | | | | | | | | | |
| Normal | –.8531 | –.8492 | –.8811 | –.8756 | –.8648[1] | .4807 | .4852 | .5028 | .4568 | .4814[1] |
| Gamma | –.8801 | –.8767 | –.8998 | –.9042 | –.8902 | .4699 | .4748 | .4922 | .4462 | .4708 |
| *Within-component variance* | | | | | | | | | | |
| .05 | –.8557 | –.8514 | –.8904 | –.8779 | –.8689 | .4827[1] | .4862 | .5011 | .4599[1] | .4825[1] |
| .25 | –.8775 | –.8745 | –.8905 | –.9018 | –.8861 | .4679 | .4738 | .4940 | .4431 | .4697 |
| *Number of households* | | | | | | | | | | |
| 75 | –.8876 | –.8830 | –.9511 | –.9031 | –.9062 | .4634[2] | .4690[2] | .4946 | .4442[2] | .4678[2] |
| 200 | –.8569 | –.8543 | –.8637[1] | –.8859 | –.8652[1] | .4767[1,2] | .4810[1,2] | .4962 | .4498[1,2] | .4759[1,2] |
| 400 | –.8553 | –.8516 | –.8565[1] | –.8807 | –.8610[1] | .4859[1] | .4900[1] | .5018 | .4604[1] | .4845[1] |
| *Number of purchases/household* | | | | | | | | | | |
| 3 | –1.7605 | –1.7539 | –1.8348 | –1.7879 | –1.7843 | .4621 | .4661 | .4945 | .4499 | .4681[2] |
| 10 | –.5035[2] | –.5009[2] | –.5024[2] | –.5253[2] | –.5080[2] | .4842[1] | .4900[1] | .5029 | .4585 | .4839[1] |
| 15 | –.3358[1] | –.3342[1] | –.3341[1] | –.3565[1] | –.3402[1] | .4796[1] | .4839[1] | .4952 | .4461 | .4762[1,2] |
| *Error variance* | | | | | | | | | | |
| Standard | –.8000[1] | –.7959[1] | –.8255[1] | –.8315[1] | –.8132[1] | .5180[1] | .5234[1] | .5420[1] | .4855[1] | .5172[1] |
| High | –.9332 | –.9301 | –.9554 | –.9483 | –.9417 | .4326 | .4366 | .4530 | .4174 | .4349 |
| Overall | –.8666[1,2] | –.8630[1] | –.8905[2] | –.8899[2] | –.8775 | .4753[2] | .4800[2] | .4975[1] | .4515[3] | .4761 |

[a]Superscripts on overall means by model type at the bottom of the table indicate significant differences (according to least significant difference rule) at the .05 level, with the superior mean having a value of "1." In the interior of the table, factor level means within model types are tested for significant differences in a similar manner. The RMSE values used in this analysis, from the ANOVAs in Table 1, are .1544 for LOGL(V) and .0633 for $\overline{P}$(V).

[b]The means for one-component data sets were not analyzed in the ANOVAs in Table 1 and therefore are not tested for significance in this table.

[c]FM-B = finite mixture models with the number of components determined by BIC; FM-L = finite mixture models with the number of components determined by LOGL(V); HB = mixed logit models estimated by hierarchical Bayes procedure; and LOGIT = homogeneous logit models.

[d]The validation log-likelihood is the posterior likelihood, computed from household-level parameter estimates. All likelihoods are divided by the number of observations in the data set to facilitate interpretation.

[e]Average choice probability of the chosen brand for the validation sample.

within-component variance is smaller, and there are more households and more purchases per household, as was expected.[8]

According to LOGL(E) (Table 2), the FM models fit significantly better than the homogeneous logit models. In general, the models fit better when the separation between components is larger (though they recover parameters better when the separation is smaller), the within-component distribution is normal rather than gamma, the within-components variance is smaller, and there is less error variance. Although the models appear to fit better when there are fewer purchases per household, this is due largely to the overfitting of the HB method for the three-purchases-per-household condition, as was discussed previously.

The parameter-adjusted fit measure –2MLL (Table 2) favors FM-B over the homogeneous logit model, which is not surprising because FM-B uses BIC as a segment retention criterion. Although HB fits significantly better than the other models according to this criterion, we again reserve judgment on HB pending further analysis of the three-purchases-per-household condition. In general, models have better parameter-adjusted fit when there are fewer components, the within-component variances are normal rather than gamma, the within-components variances are smaller rather than larger, there are more households and more purchases per household, and the error variance is smaller, as was expected.

The forecasting statistic LOGL(V) (Table 3) favors the FM-L model, in which the number of components is determined by LOGL(V). It is perhaps not surprising that FM-L performs slightly better according to this criterion because the criterion was used in the model selection process. In addition, FM-B and HB do not predict significantly better than homogeneous logit. Assuming an average data set with 2000 validation observations, the LOGL(V) advantage of FM-L over homogeneous logit would be approximately 54 likelihood points, which may or may not be important, depending on the application. In general, models forecast better according to LOGL(V) when there are fewer components, the within-component heterogeneity distributions are normal rather than gamma, there are more households and

[8]The findings for the error variance factor in Table 2 should be interpreted with caution because changes in the error variance change the scale of the parameters. In the high error variance condition, the true parameter estimates were divided by $\sqrt{2}$ before computation of RMSE($\beta$), because logit analysis implicitly assumes a scale factor of one. Thus, the RMSE($\beta$) values in the high error variance condition in Table 2 could be multiplied by $\sqrt{2}$ to facilitate comparison with the values in the standard error variance condition. Doing so shows a fairly modest increase in parameter error in the high error variance condition for most models, as we would expect.

Table 4

SIMULATION RESULTS WITH THREE-PURCHASES-PER-HOUSEHOLD CONDITION REMOVED: FM-L AND HB[a]

| | *RMSE(β)* | | *LOGL(E)* | | *–2MLL* | | *LOGL(V)* | | *$\overline{P}$(V)* | |
|---|---|---|---|---|---|---|---|---|---|---|
| *Factor* | *FM-L* | *HB* | *FM-L* | *HB* | *FM-L* | *HB* | *FM-L* | *HB* | *FM-L* | *HB* |
| *Number of components* | | | | | | | | | | |
| 1 | .3346 | .3580 | –.9674 | –.8698 | 2.0393 | 1.9250 | –.4194 | –.4207 | .4812 | .4940 |
| 2 | .4099[1] | .4181[1] | –.9567 | –.8575 | 2.0779 | 1.9172 | –.4203 | –.4212 | .4846 | .4968 |
| 3 | .4556 | .4556 | –.9332 | –.8346 | 2.0850 | 1.8832 | –.4148 | –.4153 | .4894 | .5013 |
| *Separation* | | | | | | | | | | |
| .5 | .4082[1] | .4093[1] | –.9725 | –.8713 | 2.0964 | 1.9433 | –.4240 | –.4241 | .4778 | .4908 |
| 1.0 | .4573 | .4645 | –.9175[1] | –.8208[1] | 2.0665 | 1.8572[1] | –.4111 | –.4124 | .4962[1] | .5073 |
| *Within-component distribution* | | | | | | | | | | |
| Normal | .4323 | .4396 | –.9389 | –.8413 | 2.0680 | 1.8893 | –.4134 | –.4147 | .4902 | .5021 |
| Gamma | .4332 | .4342 | –.9511 | –.8509 | 2.0949 | 1.9111 | –.4216 | –.4218 | .4837 | .4961 |
| *Within-component variance* | | | | | | | | | | |
| .05 | .3425[1] | .3843[1] | –.9397 | –.8474 | 2.0405[1] | 1.8902 | –.4105 | –.4147 | .4938 | .5034 |
| .25 | .5230 | .4894 | –.9503 | –.8448 | 2.1224 | 1.9102 | –.4246 | –.4219 | .4802 | .4947 |
| *Number of households* | | | | | | | | | | |
| 75 | .4693 | .4844 | –.9681[2] | –.8404 | 2.1600 | 1.9516[2] | –.4309[2] | –.4344[2] | .4751[2] | .4913[2] |
| 200 | .4182[1] | .4151[1] | –.9463[1,2] | –.8574 | 2.0827[2] | 1.9062[1,2] | –.4168[1,2] | –.4160[1,2] | .4854[1,2] | .4961[1,2] |
| 400 | .4107[1] | .4111[1] | –.9204[1] | –.8404 | 2.0017[1] | 1.8428[1] | –.4049[1] | –.4044[1] | .5004[1] | .5098[1] |
| *Number of purchases/household* | | | | | | | | | | |
| 10 | .4492 | .4600 | –.9299 | –.8218[1] | 2.0665 | 1.8771 | –.5009 | –.5024 | .4900 | .5029 |
| 15 | .4163[1] | .4138[1] | –.9600 | –.8703 | 2.0964 | 1.9233 | –.3342[1] | –.3341[1] | .4839 | .4952 |
| *Error variance* | | | | | | | | | | |
| Standard | .4950 | .4918 | –.8484[1] | –.7529[1] | 1.9154[1] | 1.7148[1] | –.3778[1] | –.3778[1] | .5327[1] | .5457[1] |
| High | .3705[1] | .3820[1] | –1.0415 | –.9393 | 2.2475 | 2.0856 | –.4572 | –.4587 | .4412 | .4524 |
| Overall | .4327 | .4369 | –.9450[2] | –.8461[1] | 2.0815[2] | 1.9002[1] | –.4175 | –.4183 | .4870 | .4991 |

[a]See Tables 2 and 3 for explanatory footnotes.

more purchases per household, and the error variance is smaller. Note in particular the huge forecasting benefit gained in going from three to ten estimation sample observations per household.

Finally, the other forecasting accuracy statistic, the average choice probability $\overline{P}$(V) (Table 3), favors HB. Again, we reserve judgment on this finding pending further investigation of the three-purchase condition. The FM models predict better than the homogeneous logit model according to this criterion. In general, models forecast better according to $\overline{P}$(V) when the within-component distribution of heterogeneity is normal and has smaller variance, the sample sizes are larger, and error variance is smaller.

Table 4 presents the simulation results with the three-purchases-per-household condition removed (a condition in which HB did very poorly). With this analysis, we find that HB and FM have equally good parameter estimates (RMSE[β]) and predictive capabilities (LOGL[V] and $\overline{P}$[V]) overall, though HB fits better according to LOGL(E) and –2MLL. It may initially appear surprising that HB recovers parameters as well as it does when there are three preference components in the data and the separation between components is large, because the mixed logit model implemented here does not explicitly allow for multimodal distributions. Although the prior heterogeneity distribution specified is unimodal, the posterior distribution can be multimodal.[9]

As we would expect, HB performs better than FM according to RMSE(β) when the within-component variances are larger, though the opposite is true when the within-component variances are smaller. In the interior of Table 4, differences of approximately .0182 in RMSE(β) are required for statistical significance according to the least significant difference rule. In all other experimental conditions, there are no significant differences in RMSE(β) between FM and HB.[10]

A final observation is that the in-sample log-likelihood is not a good indicator of the quality of a model, as borne out by the accuracy of the parameter estimates and predictive capability. Note in Tables 2 and 3, for example, that HB-estimated models fit the data better than do other models when the number of households and the number of purchases per household are smaller, whereas parameter recovery and predictive capability are poorer in these conditions. Thus, data analysts should be careful when selecting models on the basis of in-sample fit, even when corrections for the number of parameters are used.

## *SUMMARY, DISCUSSION, AND CONCLUSIONS*

Whereas the AAC study assesses the relative performance of metric conjoint analysis models with discrete and continuous representations of heterogeneity, this follow-up study assesses the relative performance of logit choice models with discrete and continuous representations of heterogene-

[9]This can be demonstrated by studying histograms of the posterior household parameter estimates, which are available from the authors on request.

[10]The means for one-component data are not included in the ANOVA and therefore are not tested for statistical significance.

ity using simulated scanner panel data. Given the different distributional assumptions, model forms, and model estimation procedures of metric conjoint analysis models and logit choice models and the lower information content of choice data compared with metric conjoint data, it is not at all clear that the findings of the AAC study generalize to logit choice models and scanner data.

The most important new finding of this study compared with the AAC study is the poor performance of the mixed logit model with HB estimation when parameters are poorly identified at the individual level. The AAC study did not have such a condition, so there was no opportunity to observe this finding. In general, the Bayesian literature has noted that estimation of models with few observations at the individual level is possible because the sharing of information across the full set of households is sufficient to make the model identifiable. However, our analysis clearly demonstrates that problems emerge when the number of purchases per household is too small.

When this data condition is removed from the analysis, we find that HB and FM models have equally good parameter recovery and predictive validity, though HB has an advantage in fit. The findings also show that all model specifications are quite robust to violations of underlying assumptions. These findings generalize the results of the AAC study to the context of logit choice models, despite the different distributional assumptions, model forms, and model estimation procedures of metric conjoint analysis models and logit choice models and the lower information content of choice data compared with metric conjoint data.

Whether an analyst prefers to use models with continuous or discrete representations of consumer heterogeneity is a matter of opinion and personal preference. However, empirical evidence on this issue speaks much more convincingly than subjective arguments and speculation, and we hope this research spurs other empirical work.

## REFERENCES


Abramson, Charles, Rick L. Andrews, Imran S. Currim, and Morgan Jones (2000), "Parameter Bias from Unobserved Effects in the Multinomial Logit Model of Consumer Choice," *Journal of Marketing Research*, 37 (November), 410–26.

Allenby, Greg M. and Peter E. Rossi (1999), "Marketing Models of Consumer Heterogeneity," *Journal of Econometrics*, 89 (March/April), 57–78.

Andrews, Rick L., Asim Ansari, and Imran S. Currim (2002), "Hierarchical Bayes Versus Finite Mixture Conjoint Analysis Models: A Comparison of Fit, Prediction, and Partworth Recovery," *Journal of Marketing Research*, 39 (February), 87–98.

——— and Imran S. Currim (2003), "A Comparison of Segment Retention Criteria for Finite Mixture Logit Models," *Journal of Marketing Research*, 40, forthcoming.

Brownstone, David and Kenneth Train (1999), "Forecasting New Product Penetration with Flexible Substitution Patterns," *Journal of Econometrics*, 89 (March/April), 109–29.

Chintagunta, Pradeep K., Dipak C. Jain, and Naufel J. Vilcassim (1991), "Investigating Heterogeneity in Brand Preferences in Logit Models for Panel Data," *Journal of Marketing Research*, 28 (November), 417–28.

Cohen, Jacob (1988), *Statistical Power Analysis for the Behavioral Sciences*. Hillsdale, NJ: Lawrence Erlbaum Associates.

Cutler, A. and M.P. Windham (1994), "Information-Based Validity Functionals for Mixture Analysis," in *Proceedings of the First US/Japan Conference on the Frontiers of Statistical Modeling: An Informational Approach*, Vol. 2, H. Bozdogan, ed. Boston: Kluwer Academic Publishers, 149–70.

Elrod, Terry (1988), "Choice Map: Inferring a Product Market Map from Observed Choice Behavior," *Marketing Science*, 7 (Winter), 21–40.

Erdem, Tülin (1996), "A Dynamic Analysis of Market Structure Based on Panel Data," *Marketing Science*, 15 (4), 359–78.

Gönül, Füsun and Kannan Srinivasan (1993), "Modeling Multiple Sources of Heterogeneity in Multinomial Logit Models: Methodological and Managerial Issues," *Marketing Science*, 12 (3), 213–29.

Heckman, J. and B. Singer (1984), "A Method for Minimizing the Impact of Distributional Assumptions in Econometric Models of Duration Data," *Econometrica*, 52 (March), 271–320.

Huber, Joel and Kenneth Train (2001), "On the Similarity of Classical and Bayesian Estimates of Individual Mean Partworths," *Marketing Letters*, 12 (August), 259–69.

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., Wayne S. DeSarbo, Paul E. Green, and Martin R. Young (1996), "Hierarchical Bayes Conjoint Analysis: Recovery of Partworth Heterogeneity from Reduced Experimental Designs," *Marketing Science*, 15 (2), 173–91.

McCulloch, Robert E. and Peter E. Rossi (1994), "An Exact Likelihood Approach to Analysis of the MNP Model," *Journal of Econometrics*, 64 (September/October), 207–40.

McFadden, Daniel and Kenneth Train (2000), "Mixed MNL Models for Discrete Response," *Journal of Applied Econometrics*, 15 (5), 447–70.

Newton, M.A. and A.E. Raftery (1994), "Approximate Bayesian Inference by the Weighted Likelihood Bootstrap," *Journal of the Royal Statistical Society, Series (B)*, 56 (1), 43–48.

Raftery, Adrian E. (1996), "Hypothesis Testing and Model Selection," in *Markov Chain Monte Carlo in Practice,* W.R. Gilks, S. Richardson, and D.J. Spiegelhalter, eds. London: Chapman & Hall, 163–88.

Revelt, David and Kenneth Train (1998), "Mixed Logit with Repeated Choices of Appliance Efficiency Levels," *Review of Economics and Statistics*, 80 (4), 647–57.

Rossi, Peter E., Robert E. McCulloch, and Greg M. Allenby (1996), "The Value of Purchase History Data in Target Marketing," *Marketing Science*, 15 (4), 321–40.

Schwartz, G. (1978), "Estimating the Dimension of a Model," *Annals of Statistics,* 6 (2), 461–64.

Vriens, Marco, Michel Wedel, and Tom Wilms (1996), "Metric Conjoint Segmentation Methods: A Monte Carlo Comparison," *Journal of Marketing Research*, 33 (February), 73–85.

Wedel, Michel, Wagner Kamakura, Neeraj Arora, Albert Bemmaor, Jeongwen Chiang, Terry Elrod, Rich Johnson, Peter Lenk, Scott Neslin, and Carsten Stig Poulsen (1999), "Discrete and Continuous Representations of Unobserved Heterogeneity in Choice Modeling," *Marketing Letters*, 10 (3), 219–32.