Marketing Letters 10:3 (1999): 219–232

Discrete and Continuous Representations of Unobserved Heterogeneity in Choice Modeling

MICHEL WEDEL
University of Groningen

WAGNER KAMAKURA
University of Iowa

NEERAJ ARORA
Virginia Tech

ALBERT BEMMAOR
ESSEC, Cergy-Pontoise

JEONGWEN CHIANG
Hong Kong University of Science & Technology

TERRY ELROD
University of Alberta

RICH JOHNSON
Sawtooth Software

PETER LENK
University of Michigan

SCOTT NESLIN
Dartmouth College

CARSTEN STIG POULSEN
Aalborg University

Abstract

We attempt to provide insights into how heterogeneity has been and can be addressed in choice modeling. In doing so, we deal with three topics: Models of heterogeneity, Methods of estimation and Substantive issues. In describing models we focus on discrete versus continuous representations of heterogeneity. With respect to estimation we contrast Markov Chain Monte Carlo methods and (simulated) likelihood methods. The substantive issues discussed deal with empirical tests of heterogeneity assumptions, the formation of empirical generalisations, the confounding of heterogeneity with state dependence and consideration sets, and normative segmentation.

Key words: Mixing Distributions, Multinomial Logit, Multinomial Probit, Markov-Chain Monte Carlo, Simulated Likelihood

Introduction

Unobserved heterogeneity has been widely recognized as a critical issue in modeling choice behavior, both from a theoretical and substantive standpoint (DeSarbo et al., 1997; Allenby and Rossi 1999; Wedel and Kamakura, 1997). The current state of affairs in both modeling and estimation present an opportunity to take stock of the basic ideas behind the methods involved, and to identify important debates and issues. We organize our discussion in three sections. First, we contrast continuous and discrete representations of unobserved heterogeneity. Next, we discuss the methods for obtaining individual or segment-level parameter estimates, followed by a review of the managerial issues related to consumer heterogeneity. These, in our view, represent the most important issues regarding modeling of heterogeneity in choice behavior.

1. Models of Heterogeneity: Discrete versus Continuous Distributions

The most important ways of representing heterogeneity in choice models currently in use are through either a continuous or a discrete mixture distribution of the parameters. To illustrate this, assume a model with individual-level parameters θ for $i = 1, \ldots, n$ consumers. Consider, for example, the application of a multinomial logit to scanner data. A consumer i makes a choice among J alternatives in each of T_i purchase occasions, in response to a vector X_{ijt} of predictors. The choice model is then (McFadden, 1973):

$$P(y_{ijt} = 1|\theta; X_{ijt}) = \frac{e^{\theta X_{ijt}}}{\sum_{j=1}^{J} e^{\theta X_{ijt}}}, \tag{1}$$

with $Y_{ijt} = 1$ if brand j is chosen by consumer i on occasion t and zero otherwise. For the purpose of exposition, we adopt a Bayesian framework, so that the parameters are not fixed quantities, but random variables. We are interested in the posterior distribution of the individual-level parameters, given the data. Assume that Θ is a set of (hyper) parameters indexing the distribution of individual-level parameters.

For notational simplicity, we suppress dependence on other parameters of interest. The posterior distribution of the individual-level parameters can be written as (Allenby and Rossi, 1999; Lenk and Rao, 1990):

$$\pi(\theta, \Theta|y) \propto \pi(y|\theta)\pi(\theta|\Theta)\pi(\Theta), \tag{2}$$

where the three terms after the proportionality sign are the likelihood, the mixing distribution and the prior for Θ, respectively. For example, in the MNL the likelihood for consumer i (assuming independence across purchase occasions) is given by

$$\pi_i(y|\theta) = \prod_{t=1}^{T_1} \prod_{j=1}^{J} P(y_{ijt} = 1|\theta, X_{it})^{y_{ijt}}, \tag{3}$$

while the mixing distribution could be the multivariate normal: $\pi(\theta|\Theta) = \mathrm{MVN}(\mu, \Sigma)$. Frequentist inference, which does not take prior information on the parameters into account (i.e. $\pi(\Theta)$ is ommited), focuses on obtaining point estimates of the (hyper-) parameters given the observed data (the likelihood does not involve a probability measure on the parameters, cf. Lindsey 1996, p. 76). Equation (2) then involves an integration over the mixing distribution: $\pi(\Theta|y) = \int \pi(y|\theta)\pi(\theta|\Theta)d\theta$ The discussion below focuses on the form of the mixing distribution, $\pi(\theta|\Theta)^2$.

Aggregate models of choice (e.g. Guadagni and Little, 1983) are the simplest form of these as they attempt to model choice assuming a homogeneous population. In the aggregate models the choice parameters θ (e.g. price sensitivity) do not vary across the population and there is no mixing distribution. However, even those models accommodate heterogeneity because individual preferences are incorporated as independent variables, which vary across households. Individual-level predictions can be made with those models and they may even predict well, but on a strict sense, this is not a model of heterogeneity.

Early approaches to heterogeneity treated heterogeneity as a nuisance, and included individual-level intercept terms into the choice model to eliminate heterogeneity. First, so called fixed effects approaches were used where individual-level parameters were included in the model and could be estimated directly. Later, conditional likelihood approaches were used, in which the model was formulated conditional upon sufficient statistics for the individual-level parameters, which eliminated heterogeneity effects from the model and greatly simplified the estimation task (Chamberlain, 1980). Subsequently an (unspecified) distribution was assumed for the intercept term. This assumed continuous distribution was approximated by a discrete number of support points and probability masses (Heckman and Singer, 1984; Chintagunta, Jain and Vilcassim, 1991), which involves the mixing distribution as defined in equation $1'(2){:}\pi(\theta/\Theta) = \pi_s$, for $s = 1, \ldots, S$ Later, heterogeneity became of fundamental interest itself and it was noted that heterogeneity pertained potentially to all the parameters in a model. Thus the support point approach was extended to capture heterogeneity across all the parameters in a choice model.

Thus, finite mixture regression models arose that connected very well to marketing theories of market segmentation (Wedel and Kamakura, 1997). Such finite mixture models (Kamakura and Russell, 1989; DeSarbo, Ramaswamy and Cohen, 1995) have received considerable attention by practitioners and academics. Managers seem comfortable with the idea of market segments, and the models appear to do a good job of identifying useful groups. However, market segments cannot account fully for heterogeneity in preference if the underlying distribution of preference is in fact continuous. Many practitioners, such as direct and database marketers, prefer to work at the level of the individual respondent. The assumption of within-group homogeneity is often ignored in making predictions, and individual estimates are obtained as weighted combinations of segment-level estimates, where the weights are the posterior probabilities of segment membership. This is an empirical Bayes method of obtaining individual-level estimates.

While a discrete mixing distribution leads to finite mixture models, continuous mixing distributions lead to random coefficients (e.g. probit or logit) models. Random coefficient logit models are also called mixed logit models, and the models have also been called hierarchical or multi-level models. Such models have received considerable attention in marketing and related fields (cf. Allenby and Ginter, 1995; Allenby and Lenk, 1994; Rossi, McCulloch and Allenby, 1996; Train and Brownstone, 1998; Elrod and Keane 1995; Haaijer, Wedel, Vriens and Wansbeek, 1998; Haaijer, Kamakura and Wedel, 1998). Typically a multivariate normal distribution is assumed for all regression parameters in the model, $\pi(\theta|\Theta) = \text{MVN}\ (\mu, \Sigma)$ but other distributions can be assumed. Continuous distributions have several advantages: they seem to characterise the tails of the heterogeneity distribution better and predict individual choice behavior more accurately than finite mixture models (Allenby and Rossi, 1999; Allenby, Arora and Ginter, 1998). They provide flexibility with regard to the appropriate (in terms of the parameter space) choice of the distribution of heterogeneity. This allows model specification to closely follow an underlying theory of consumer behavior (see for example Allenby, Arora and Ginter, 1998). Moreover, individual level estimates of model parameters are easily obtained. Rossi and Allenby (1993) offer a fixed-effect (Bayesian) model, where information other than a panel members data is used to estimate the parameters, and this information can come from other panel members (through a mixing distribution) or through a prior that sets reasonable bounds on the parameters.

Critique levied against the discrete mixture approach to heterogeneity is that its predictive power in hold-out samples of alternatives is limited because individual-level estimates are constrained to lie in the convex hull of the class-level estimates. Because of this, models at the individual-level, or models with continuous heterogeneity distributions have been found to outperform the mixture model approaches (Vriens, Wedel and Wilms 1995; Lenk, DeSarbo, Green and Young, 1996).

Practical solutions to the convex hull problem have been proposed in the conjoint choice framework. Johnson (1997) proposed a model that involves individual part worths from conjoint choice data, named “ICE” (Individual Choice Estimation), which uses a lower-rank approximation to the subjects $\times$ variables matrix of individual partworths. ICE is similar to a method proposed by Hagerty (1985), although Hagerty’s model dealt with OLS estimation whereas ICE uses Logit estimation. ICE finds estimates of individual partworths, which lie in that subspace but which are not confined to the convex hull of the segment-level parameters. ICE seems to hold promise from a practitioner point of view, but currently some unresolved problems of model identification surround it and need further study.

Some have argued that the underlying assumption of a limited number of segments of individuals that are perfectly homogeneous within segments in finite mixture models is overly restrictive (cf. Allenby and Rossi, 1998). To those authors, market segmentation in choice modeling leads to an artificial partition of the continuous distribution into homogeneous segments. If the underlying distribution is continuous, then assuming a discrete mixing distribution leads to inconsistent parameter estimates. It has been argued that in many fields within marketing, emphasis is now on individual customer contact and direct marketing approaches, and that individual-level response parameters are required for optimal implementation of direct and micro marketing strategies.

On the other hand, proponents of the discrete heterogeneity approach have put forward the proposition that the estimates of models with continuous heterogeneity distributions may be sensitive to the specific distribution assumed for the parameters (i.e. the normal), which is defined subjectively by the researcher. Further, most models that approximate heterogeneity through a number of unobserved segments have great managerial appeal: some companies have started to scale down product assortment to target larger segments with more limited variety of products. Models including segment-level estimates have an edge when there are scale advantages in production, distribution, or advertising. In addition, several approaches have been developed that allow for within-segment heterogeneity by compounding the distribution for the dependent variable, for example a Multinomial distribution for y, with a conjugate heterogeneity distribution, such as the Dirichlet, giving rise to the Dirichlet-Multinomial, that effectively captures over-dispersion of the dependent variable within classes (see for example, Böckenholt, 1993).

To a large extent, the issue of a continuous versus a discrete distribution of heterogeneity is an empirical one. A continuous heterogeneity distribution can be approximated closely by a discrete one by letting the number of support points of the discrete distribution increase at the cost of a decrease in the reliability of the parameters. For some products and markets the assumption of a number of homogeneous underlying segments may be tenable while in other cases a continuous heterogeneity distribution may be more appropriate. Whether the estimation results are managerially actionable also plays a role in selecting the appropriate model. For some applications, managers can only address a finite number of relatively homogeneous market segments, while in others (such as direct marketing), managers might be more interested in individual-level estimates for each of their customers. We need more simulation studies (e.g., Vriens, Wedel Wilms, 1995) and more empirical studies (e.g., Lenk et al., 1996) to fully understand the strengths and weaknesses of the several methods for estimating individual-level parameters. More recently, combinations of the discrete and continuous heterogeneity approaches have been developed, that account for both discrete segments and within segment heterogeneity (Allenby, Arora and Ginter, 1998; Allenby and Rossi, 1998; Lenk and DeSarbo, 1998). Table 1 summarizes several of the issues discussed above.

2. Methods of Estimation: Markov Chain Monte Carlo versus Simulated Likelihood

The two most important ways in which choice models with heterogeneity have been estimated is through maximizing a likelihood function, and with Bayesian approaches. Both discrete and continuous heterogeneity models can in principle be estimated with either maximum likelihood or Bayesian methods: in fact all models in Table 1 can be estimated with both approaches. However, currently many published papers in marketing that utilize a continuous distribution of heterogeneity have relied upon Bayesian methods. The advantage of Bayesian methods lies in obtaining posterior distributions of individual-level parameters, based on the actual distribution of the hyperparameters is used. ML methods approximate the posterior distribution of hyperparameters by quadratic approximations to the likelihood around the point estimates, and posterior estimates of individual level parameters can only be obtained by using empirical Bayes estimates, conditioning on the point estimates of the hyperparameters.

Table 1. Model Comparison

Mixing Distribution	Aggregate	Discrete	Continuous	Continuous & Discrete
Example reference	Guadagni & Little (1983)	Kamakura & Russell (1989)	Allenby & Lenk (1995)	Allenby, Arora & Ginter (1998)
Typical model name	Multinomial Logit	Latent Class Logit or Mixture Logit	Multinomial Probit, Random Coefficients Probit, or Mixed Logit	Mixture Multinomial Probit
Most used Estimation method[1]	ML	ML	SML, MCMC, IS	MCMC
Individual-level predictions	Yes, but dependent on individual-level predictor variables.	Yes, but constrained in a convex hull	Yes, but influenced by aggregate parameters	Yes, but influenced by component parameters
Precision of individual-level estimates	Not applicable	Not available	Obtained empirically from iterates of a MCMC or IS only	Obtained empirically from iterates of the MCMC
Functions of model parameters	Delta method approximation	Delta method approximation	ML: Delta method approximation, for MCMC or IS obtained empirically from iterates	Obtained empirically from MCMC or IS iterates
Heterogeneity distributions accommodated[2,3]	Compound distributions for y, e.g. BB, DM	M for θ and compound for y e.g. BB, DM	Arbitrary continuous for θ, e.g. N, G, TN	M and arbitrary continuous for θ
Nested models	–	Aggregate	Aggregate	Aggregate, Discrete and Continuous

[1] ML = Maximum Likelihood, SML = Simulated ML, MCMC = Markov Chain Monte Carlo, IS = Geweke's Importance Sampling.
[2] BB = Beta-Binomial, G = Gamma, DM = Dirichlet-Multinomial, N = Normal, M = Multinomial TN = Truncated Normal.
[3]y and θ are defined in the text.

Bayesian estimation methods have gained popularity recently because they provide a set of techniques that allow for the development and analysis of complex models. The widely used Monte Carlo Markov Chain (MCMC, e.g. Gelman et al., 1995) methods involve integration over the posterior distribution of the parameters given the data (equation 2) by drawing samples from that distribution. Starting from equation (2), this would involve successively drawing samples from the full conditional distributions of the model parameters (Allenby and Lenk, 1994). Many applications in marketing involve the Gibbs-sampler as a special case, which can be implemented if expressions for the full conditional distributions of all parameters can be obtained. If that is not the case, powerful alternatives such as the Metropolis-Hastings algorithm are available that, based on a known candidate distribution, involve a rejection-type of sampling method to approximate those posterior distributions. Sample statistics such as the mean, mode and other percentiles, are then computed from the draws to characterize the posterior distribution. Statistical properties of estimates (e.g. precision) and estimates of functions of model parameters are thus easily obtained empirically. The latter feature is particularly useful in the evaluation of non-linear functions of model parameters since for non-linear functions $(f(\theta), E(F(\theta)) \neq f(E(\theta))$. As marketing researchers move in the direction of utilizing heterogeneity models to make decisions, this feature of hierarchical Bayes models becomes an important advantage. However, there are two concerns with regard to the use of hierarchical Bayes model. First, the distribution used to characterize heterogeneity is determined subjectively by the researcher. Second, the simulation-based estimation procedure such as Gibbs sampling may be more computer intensive than the maximum likelihood approach. The second problem is becoming less of an issue because of the availability of faster computers. A pragmatic fix for the first problem is a sensitivity analysis with regard to the choice of distribution in order to check model robustness (for model checking procedures see Allenby and Rossi, 1999). From a more dogmatic Bayesian point of view, the subjective choice of the distribution characterizes the analysts' uncertain state of knowledge, which does not need to conform to that of other analysts in this matter.

Although recent advances in MCMC may provide pragmatic Bayesians with a slight edge in estimating complex models, non-Bayesian methods, such as EM or the method of simulated maximum likelihood discussed below, are rapidly closing the performance gap. Will the rush to adopt Bayesian methods be followed by a bust as these alternatives assert themselves? The more dogmatic Bayesian think not because of the rich philosophical foundation of Bayesian inference (cf. Bernardo and Smith, 1994; De Finetti, 1970; Jeffreys, 1939; and Savage, 1954). What starts as an expedient to obtain a solution often ends in transforming the user. Berger's experience is not atypical. In the preface to the first edition of his book (Berger, 1980, page vii), he describes his gradual conversion in a remarkable moment of candor:

"Specific considerations that I found particularly compelling were: (i) The Bayesian measures of accuracy of a conclusion seem more realistic than the classical measures . . . . (ii) In most circumstances any reasonable statistical procedure corresponds to a Bayes procedure . . . . (iii) Principles of rational behavior seem to imply that one must act as if he had a prior distribution."

But there may be reservations about Bayesian philosophy, which parallel those of Berger (1980) involving sensitivity to the specification of the prior, and the lack of objectivity. The first reservation can be overcome by performing a sensitivity study of the results to the prior specification. At the very least, Bayesians are explicit about their prior assumptions for all to judge. The loss of objectivity from the scientific method is difficult to overcome because the ideal of "objectivity" is ingrained into our education as scientific researchers. Despite the veneer of objectivity, researchers make numerous subjective choices in the selection of problems, the collection of data, the choice of models, and the presentation and interpretation of results. The debate between the objective and subjective camps focuses on the rather narrow issue of parameter estimation: should *a priori* beliefs be allowed to affect the parameter estimates, or should these estimates be strictly a function of the data and model?

Subjective Bayesians propose that a model reflects a researcher's belief about a phenomenon and is designed as an aid in directing his or her thinking about various aspects of that phenomenon. But why publish researchers' subjective findings that express their internal states of knowledge and ignorance and need not be linked to a "true" model? At the extreme, subjectivism seems to reflect an "anything goes" mentality, which is in opposition to the rigors of scientific discipline. A more moderate view is that the informed opinions of researchers advance science by modifying the beliefs of others. The more mainstream applied Bayesian believes that the phenomenon has a true, objective model, which can be revealed through a researcher's investigations. This hybrid approach often attempts to reconcile Bayesian methods with "objectivity" by using an objective likelihood function and subjective prior distributions. This approach, however, ignores vital issues such as model uncertainty and model verification (Lenk, 1998). To fully reap the benefits of Bayesian inference, a pragmatic Bayesian needs to jump the abyss from objectivity to subjectivity. Not to do so ultimately leaves the pragmatic Bayesian in the unpleasant position of defending what he or she personally believes to be indefensible.

Other estimation methods avoid the stumbling blocks of the objective/subjective controversy by focusing solely on the likelihood function. The method of maximum likelihood, the dominant frequentist approach to estimating choice models with heterogeneity, has several variants, including the EM and Stochastic EM. The EM algorithm iterates between two steps: the E-step which involves taking the expectation over the mixing distribution of the unobserved heterogeneity parameters and the M step which maximizes the Expected likelihood obtained in the E-step over the remaining parameters. In the SEM algorithm (Diebolt and Ip, 1996) instead of taking the expectation in the E-step, a draw from the distribution of the heterogeneity parameters is taken. A third method for the estimation of choice models with heterogeneity is the method of Simulated Maximum Likelihood, which has received considerable attention in econometrics (Gouriéroux and Monfort, 1993; Lee, 1995, 1997; Revelt and Train, 1997) and has some conceptual similarities with SEM. As compared to MCMC, for which the distribution of nonlinear functions of the parameters are obtained empirically from the iterates (cf. Table 1), their precision in a maximum likelihood framework can be obtained through the delta method, involving a quadratic approximation that is approximately valid as the sample size tends to infinity (Lindsey, 1996, p. 205).

Consider again the multinomial logit model. For a continuous mixing distribution (Revelt and Train, 1997), the log-likelihood for consumer i is given by

$$\ell_i(y|\Theta) = \ln\left(\int \prod_{t=1}^{T_i} \prod_{j=1}^{J} P(y_{ijt} = 1|\theta, X_{ijt})^{y_{ijt}}\ \pi(\theta|\Theta)d\theta\right) \tag{4}$$

The integration is not tractable in most applications. In order to circumvent this problem, the method of Simulated Maximum Likelihood (SML) requires R draws from the mixing distribution $\pi(\theta|\Theta)$, and the approximation of the integral by an average computed over these random draws. The GHK simulator (*G*eweke-*H*ajivassiliou-K*eane*) is a preferred choice in the SML context for this purpose (e.g., Elrod and Keane, 1995; Börsch-Supan and Hajivassiliou, 1993; Hajivassiliou, 1993). The simulated log-likelihood is asymptotically unbiased, where the bias decreases proportionally to $\sqrt{n}/r$. For samples of $n = 400$ and $R = 50$ replications the method has been shown to provide good performance on synthetic data (see Lee, 1995; Lee, 1997), but that performance may still critically depend on the dimension of the integration involved. A very appealing aspect of the SML estimator is that the simulated likelihood is twice differentiable, simplifying its implementation with gradient search algorithms. It also provides individual-level estimates of the response coefficients. Parsimonious accounts of the unobserved heterogeneity can be attained by imposing a factor structure on the covariance of the random coefficients (Gönül and Srinivasan, 1993; Haaijer, Wedel, Vriens and Wansbeek, 1998).

Under certain conditions (Lindsey, 1996, p. 336) the ML and Bayesian approach lead to the same results. For example, for large n the two approaches converge. In that case the posterior distribution approximates the normal with a covariance matrix which is equal to the inverse of the Hessian evaluated at the maximum likelihood estimates. So, the approaches are equivalent for practical purposes if the database is large, thus providing a pragmatic motivation for continued use of frequentist (ML-based) methods under conditions that occur for many marketing applications. However, in particular for small samples and certain parameterizations, the Bayesian approach, involving MCMC estimation, provides much more accurate approximations of the posterior distribution of the parameters.

Geweke (1989) lays out a procedure for obtaining “pure” Bayes posteriors from an empirical Bayes model estimated by maximum likelihood, i.e. starting from the MLE and the Hessian. The method is an application of importance sampling. The essence is that for some function $g(\Theta)$ on the real domain and a known distribution $h(\Theta)$

$$E[g(\Theta)] = \int g(\Theta)\frac{\pi(\Theta|y)}{h(\Theta)}d\Theta \tag{6}$$

By drawing R simulates from $h(\Theta)$ one obtains an estimator of $E[g(\Theta)]$:

$$E[g(\Theta)] = \frac{\sum_r g(\Theta_r)w(\Theta_r)/r}{\sum_r w(\Theta_r)/R}, \tag{7}$$

where $w(\Theta_r) = \pi(\Theta_r|y)/h(\Theta_r)$ and $\pi(\Theta_r|y)$ is unknown but proportional to the likelihood. A distribution such as the Multivariate-t, with it fist moment equal to the MLE and second moment proportional to the Hessian may be used as an initial approximation in a weighted sampling procedure. The distribution used to generate the simulates is updated automatically based upon initial draws. The simulates are used to characterize the posterior distributions in much the same way as for Markov chain Monte Carlo simulates. In Geweke's procedure, discrepancies between the distribution used to generate the simulates and the true distribution are corrected for by the weights. However SML methods need to be used if the model includes heterogeneity. The procedure explicitly bridges the gap between MCMC and SML approaches to estimating choice models with heterogeneity.

3. Substantive Issues

In order to model consumer choice behavior, one needs to make several assumptions regarding heterogeneity, the appropriateness of which is largely an empirical issue. Therefore, one should empirically assess the relative contribution of potential sources and formulations of heterogeneity, through nested model tests and investigation of predictive validity. Examples are provided by Allenby and Lenk (1994), and Allenby, Arora and Ginter (1997). Such model tests will ultimately allow for empirical generalization of heterogeneity findings, which can be used in attempting to answer questions such as: *In which conditions is a continuous type of heterogeneity or a discrete one more appropriate? In which conditions do what consumer descriptor variables adequately capture heterogeneity? What is the "optimal" number of segments in finite mixtures? What distribution should be assumed if heterogeneity is continuous? What is the effect of allowing for within segment heterogeneity on the number of segments in mixture models?* Such empirical generalizations could help to form theoretical foundations for the description of heterogeneity.

An important issue for future research is to provide such a theoretical underpinning of heterogeneity, with the purpose of identifying variables that need to be included in models and to assist researchers in the appropriate model specification. Future models should be based on verifiable assumptions about the underlying process that generates heterogeneity. One of the most challenging aspects of modeling heterogeneity stems from its link to state dependence, or purchase event feedback (Allenby and Lenk, 1995). Purchase event feedback refers to the impact of current choices on future choices. Purchase event feedback is a dynamic concept, whereas heterogeneity is a static, cross-sectional concept. However, there is an inherent relationship between feedback and heterogeneity because consumers accumulate different amounts of purchase event feedback over time, and this in turn gives rise to a revised heterogeneity at any point in time. The challenge is to disentangle these two effects. An important question is how to include both heterogeneity and purchase event feedback in a way that captures each phenomenon while allowing for an unbiased partitioning of dynamic feedback and constant cross-sectional heterogeneity. Keane (1997) proposes an extensive array of nested models and tests to disentangle those effects. Chiang, Chib and Narasimhan (1997) focus on the distinction between heterogeneity in the response parameters and heterogeneity in the consideration set. They argue that it is important to incorporate both forms of heterogeneity in choice models.

Importance lies in the association of a unique set of managerial implications for each source of heterogeneity. Chiang and co-authors show one way of tackling the complexity of estimating both components in a choice model. The model is a random coefficients model in which choice set heterogeneity is accommodated by considering the power set of all possible brands. MCMC methods are applied to obtain the posterior distributions of the parameters in the choice model as well as of the consideration sets. Nested model tests show the importance of accounting for both heterogeneity and consideration sets. A final theme for future research is to revisit segmentation from a normative approach. There are few contributions in marketing to that topic, Mahajan and Jain (1978) being one exception. Revenues and costs need to be included in the segmentation model. So far, only statistical 'costs' have been incorporated, and that is clearly not satisfactory and leaves us with unresolved problems like determining the 'right' number of segments. Advantages of scale in production and marketing are crucial, since without such advantages of scale, an individual approach, as advocated in direct and micro-marketing seems appropriate. Therefore, segmentation should be viewed as a decision problem rather than a mere statistical problem, and normative segmentation should be based on a microeconomic foundation. A practical approach could be to include some elements of the decision problem into a loss-function to be minimized. Bayesian approaches appear to have merit in this respect.

4. Conclusion

The advantage of a discrete heterogeneity distribution is that it doesn't rely on a parametric form that may be inaccurate. Also, the market segments from this model are often very compelling from a managerial standpoint. Its disadvantage is that it can over-simplify and that of limited predictive validity. Continuous representations are capable in theory of capturing the true distribution because the true distribution may often be continuous. In addition, it does not impose overly restrictive constraints on individual parameters. The problem with continuous representations is that the well-behaved parametric distributions we find easy to use may not be flexible enough to capture the true distribution. Both the simulated likelihood and the Markov Chain Monte Carlo methods can be used in principle to estimate choice models with heterogeneity. Although from a philosophical standpoint the approaches are distinct, from a pragmatic standpoint the procedures converge, in particular for non-informative priors and large samples. Geweke's procedure further bridges the gap between the two classes of methods. However, advantages of the Bayesian approach accrue in particular when one wants to obtain posterior distributions of individual-level parameters, and when one takes the stance that *a priori* beliefs should be allowed to affect the parameter estimates. More empirical studies are needed to assess how serious the weaknesses of models and estimation methods are in numerical terms, how models and estimation procedures compare across a wide range of conditions and to provide guidelines for which particular substantive problem a particular method is most suited. An important consideration is whether one is interested in predicting the future behavior (such as in direct marketing data for financial products) or in obtaining results that are representative for the total market or population, based on a *sample* of individuals (as household-level scanner data on non-durable goods). Models with continuous representation of heterogeneity appear to be doing better than those with discrete heterogeneity in the former case (especially when estimated with Bayesian methods that allow for individual level estimates to be obtained). However, the latter issue is unresolved for those methods, since individual level estimates can only be obtained for individuals in the sample. In addition, there are a number of key substantive issues that are relevant in modeling heterogeneity, the most important of which include heterogeneity and state dependence, choice set heterogeneity, better theoretical foundation of the existence of heterogeneity, and approaches to optimize economic rather than statistical criteria.

Note

1. The density function $\pi(\cdot)$ is with respect to arbitrary measure so that $\pi(\cdot)$ can be a discrete mass function by adopting a counting measure.

References

Allenby, Greg M., Neeraj Arora and James L. Ginter (1998), "On the Heterogeneity of Demand," *Journal of Marketing Research*, August 35, 384–389.

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

Allenby, Greg M. and Peter J. Lenk (1994), "Modeling Household Purchase Behavior with Logistic Normal Regression," *Journal of the American Statistical Association*, December, Vol. 89, No. 428, 1218–1231.

Allenby, Greg M. and Peter J. Lenk (1995), "Reassessing Brand Loyalty, Price Sensitivity, And Merchandising Effects On Consumer Brand Choice," *Journal of Business and Economic Statistics*, 13 (3), 281–290.

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

Arora, Neeraj, Greg M. Allenby and James L. Ginter (1998), "A Hierarchical Bayes Model of Primary and Secondary Demand," *Marketing Science*, Vol. 17, No. 1, 29–44.

Berger, James O., *Statistical Decision Theory: Foundations, Concepts and Methods*, First Edition, Springer-Verlag, New York, 1980.

Bernardo, Jose M. and Adrian F. M. Smith, *Bayesian Theory*, John Wiley & Sons, New York, 1994.

Böckenholt, Ulf (1993). Estimating Latent Distributions in Recurrent Choice Data, *Psychometrika*, 58, 489–509.

Börsch-Supan, Axel and Vassilis A. Hajivassiliou (1993), "Smooth Unbiased Multivariate Probability Simulators for Maximum Likelihood Estimation of Limited Dependent Variable Models," *Journal of Econometrics*, 58(3), 347–368.

Chamberlain, Gary (1980) "Analysis of Covariance with Qualitative Data," *Review of Economic Studies*, 46, 225–238

Chiang, Jeongwen, Siddarth Chib and C. Narasimhan (1999), "Markov Chain Monte Carlo and Models of Consideration Set and Parameter Heterogeneity," *Journal of Econometrics*, 89, 223–248.

Chib, Siddarth and E. Greenberg (1995), "Understanding the Metropolis-Hastings Algorithm," *The American Statistician* 49, 327–335.

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

Davidian, Marie and Gallant, A. R. (1993), "The nonlinear mixed effects model with a smooth random effects density," *Biometrika*, 80, 475–488.

Davidian, Marie and David M. Giltinan (1995), *Nonlinear Models for Repeated Measurement Data*, London: Chapman & Hall.

De Finetti, Bruno, *Teoria Delle Probabilita*, Giulio Einaudi editore s.p.a., Torino, 1970. Translated by Antonio Machi and Adrian Smith and published in two volumes by John Wiley & Sons, New York, under the title *Theory of Probability*, 1974 and 1975.

DeSarbo, Wayne S., Venkat Ramaswamy, and Steve H. Cohen (1995), "Market Segmentation with Choice-Based Conjoint Analysis," *Marketing Letters*, 6, 137–148.

DeSarbo, Wayne S., Asim Ansari, Pradeep Chintagunta, Charles Himmelberg, Kamel Jedidi, Rich Johnson, Wagner A. Kamakura, Peter Lenk, Kannan Srinivasan, Michel Wedel (1997), Representing Heterogeneity in Consumer response models". *Marketing Letters*, 8(3) 335–348.

Diebolt, Jean and Eddie H.S. Ip (1996), Stochastic EM: Method and Application. In: Markov Chain Monte Carlo in Practice (W.R. Gilks, S. Ridchardson and D.J. Spiegelhalter eds.) Chapman and Hall, London, pp. 259–274.

Elrod, Terry and Michael Keane (1996), "A Factor Analytic Probit Model for Representing the Market Structure in Panel Data," *Journal of Marketing Research*, 32 (February), 1–16.

Fader, Peter S., and James M. Lattin (1993), "Accounting for Heterogeneity and Nonstationarity in a Cross-Sectional Model of Consumer Purchase Behavior," *Marketing Science* 12, 304–317.

Gelman, Andrew, John B. Carlin, Hal S. Stern and Donald B. Rubin (1995), *Bayesian Data Analysis*. London: Chapman Hall.

Geweke, John (1989), "Bayesian Inference in Econometric Models Using Monte Carlo Integration," *Econometrica*, 57, 1317–1339.

Gönül, Füsun and Kannan Srinivasan (1993), "Modeling Multiple Sources of Heterogeneity in Multinomial Logit Models: Methodological and Managerial Issues," *Marketing Science* 12, (3), 213–229.

Gouriéroux, Christian and Alain Monfort (1993), "Simulation-based inference," *Journal of Econometrics*, 59, 5–33.

Guadagni, Peter M. and John. D.C. Little (1983), "A Logit Model of Brand Choice," *Marketing Science*, 2 (Summer), 203–238.

Haaijer, Rinus E., Michel Wedel, Marco Vriens and Tom J. Wansbeek (1998) Predictions in conjoint choice experiments: The X-factor probit model. *Marketing Science*, 17(3), 236–253.

Hagerty, Michael R. (1985), "Improving Predictive Power of Conjoint Analysis: The Use of Factor Analysis and Cluster Analysis," *Journal of Marketing Research*, 22, May, 168–184.

Hajivassiliou, Vassilis A. (1993), "Simulation Estimation Methods for Limited Dependent Variable Models," in G.S. Maddala, C.R. Rao, and H.D. Vinod (eds), *Handbook of Statistics*, Vol 11.

Jeffreys, Harold, *Theory of Probability*, first published by International Series of Monographs in Physics, 1939, reprinted by Oxford University Press, 1985.

Johnson, Richard M. (1997), "ICE: Individual Choice Estimation" Working paper, Sawtooth Software, Inc.

Kamakura, Wagner A. and Gary J. Russell. (1989), "A Probabilistic Choice Model for Market Segmentation and Elasticity Structure." *Journal of Marketing Research*, 26, 339–390.

Keane, Michael P. (1997), "Modeling Heterogeneity and State Dependence in Consumer Choice Behavior," *Journal of Business and Economic Statistics*, 15 (July), 310–327.

Lee, Lung-Fei (1995), "Asymptotic Bias in Simulated Maximum Likelihood Estimation of Discrete Choice Models," *Econometric Theory*, 437–483.

Lee, Lung-Fei (1997), "Simulated Maximum Likelihood Estimation of Dynamic Discrete Choice Statistical Models: Some Monte Carlo results," *Journal of Econometrics*, 82, 1–35.

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*, Vol. 15, No. 2, 173–191.

Lenk, P. (1998), "Bayesian inference of semiparametric regression," *Journal of the Royal Statistical Society*, Series B, forthcoming

Lenk, P. and DeSarbo, W. (1998), “Bayesian inference for finite mixtures of generalized linear models with random effects,” *Psychometrika*, forthcoming

Lindsey, Jim K. (1996), *Parametric Statistical Inference*. Oxford: Clarendon Press.

McFadden, D. (1973), “Conditional Logit Analysis of Qualitative Choice Behavior,” In P. Zarembka (ed.), *Frontiers in Econometrics*, New York: Academic.

Mahajan, Vijay and Arun K. Jain. “An Approach to Normative Segmentation,” *Journal of Marketing Research*, August 1978, 338–345

Revelt, David and Kenneth Train (1997), “Mixed Logit with Repeated Choices: Household's Choices of Appliance Efficiency Level,” *Review of Economics and Statistics* (Forthcoming)

Rossi, Peter and Greg Allenby (1993), “A Bayesian Approach to Estimating Household Parameters,” *Journal of Marketing Research* 30, 171–82.

Rossi, Peter E., Robert E. McCulloch and Greg M. Allenby (1996), “The Value of Purchase History Data in Target Marketing,” *Marketing Science*, 15, 321–340.

Savage, Leonard J., *The Foundations of Statistics*, 1954, reprinted by Dover, New York, 1972.

Vriens, Marco, Michel Wedel, and Tom Wilms. “Segmentation methods for metric conjoint analysis: A Monte Carlo comparison,” *Journal of Marketing Research*, 33 (February), 1996, 73–85,

Wedel, Michel and Wagner A. Kamakura (1997), *Market Segmentation: Conceptual and Methodological Foundations*. Dordrecht: Kluwer.