Marketing Letters 8:3 (1997): 335–348

Representing Heterogeneity in Consumer Response Models

1996 Choice Conference Participants

WAYNE S. DESARBO (CHAIR)
Pennsylvania State University, Department of Marketing, College of Business Administration, 701 Business Adminstration Building, University Park, PA 16802-3007

ASIM ANSARI
Columbia University

PRADEEP CHINTAGUNTA
University of Chicago

CHARLES HIMMELBERG
Columbia University

KAMEL JEDIDI
Columbia University

RICHARD JOHNSON
Sawtooth Software

WAGNER KAMAKURA
University of Pittsburgh

PETER LENK
University of Michigan

KANNAN SRINIVASAN
Carnegie Mellon University

MICHEL WEDEL
University of Groningen

Abstract

We define sources of heterogeneity in consumer utility functions related to individual differences in response tendencies, drivers of utility, form of the consumer utility function, perceptions of attributes, state dependencies, and stochasticity. A variety of alternative modeling approaches are reviewed that accommodate subsets of these various sources including clusterwise regression, latent structure models, compound distributions, random coefficients models, etc. We conclude by defining a number of promising research areas in this field.

Key words: Heterogeneity, latent structure models, clusterwise regression, random coefficients models, compound distributions

1. Introduction

We explore how heterogeneity affects the specification and estimation of various choice response models developed and used in Marketing. As to be discussed, heterogeneity is a result of the individual differences consumers evince with respect to the judgments they make and the processes involved in making such judgments. For decades, Marketers have debated over the value of estimating such response models at a variety of alternative levels of aggregation to perhaps gain additional insight into this heterogeneity.

1.1. Level of aggregation

Modeling approaches in Marketing have typically been implemented at one of three levels of data aggregation: aggregate, market segment, and individual consumer levels. At the *aggregate level*, one response function is estimated pooling across the data collected for the entire sample. The parameters of this aggregate response function may or may not be explicitly posited to have a distribution across the population of study. Predictions outside of the sample are typically made with this single set of parameter values, and thus may not fully capture individual consumer differences in the sample. In modeling at the *market segment level*, response parameters are typically estimated per market segment (in either apriori or post hoc segmentation schemes). In many of the post hoc approaches, the size and structure of the market segmentation scheme are also estimated simultaneously. The assumption here is that heterogeneity is adequately captured by discrete classes or groups. Predictions for individual members are conditioned upon their segment membership. Finally, at the *individual consumer level*, response parameters are estimated for each individual consumer separately. For example, conjoint analysis studies often involve individual level estimation from which market choice simulators are calibrated. This approach typically involves the collection of rather detailed replicated data per consumer as well as the estimation of a potentially huge number of parameters which may jeopardize degrees of freedom and parameter stability. Yet, individual level estimation can provide the maximum amount of flexibility in appropriately modeling consumer heterogeneity. Obviously, the selection of which level of aggregation is most appropriate is an empirical issue depending on the nature and form of consumer heterogeneity, which we next discuss.

1.2. The nature of heterogeneity

To illustrate this issue, assume the following general utility function (choice probability, intention to buy, preference, etc.):

$$Y_{ijt} = F_{it}[X_{ijkt}; B_{0it}, \mathbf{B}_{it}] + e_{ijt} \qquad (1)$$

where:

i	=	1,…, I consumers;
j	=	1,…, J brands;
k	=	1,…, K brand attributes;
s	=	1,…, S market segments;
t	=	1,…, T time periods or purchase occasions;
Y_{ijt}	=	utility for brand j in period t by consumer i;
B_{0it}	=	intercept;
$\mathbf{B}_{it}$	=	response parameters for the K brand attributes;
X_{ijkt}	=	the value of the k-th attribute for the j-th brand evaluated by consumer i in the t-th time period;
F_{it}	=	the functional form utilized by consumer i in period t;
e_{ijt}	=	error distributed as $g(\theta)$.

Note that utility, Y_{ijt}, may be directly observed as with preference ratings, or may be latent and unobserved as in the case of stated or revealed choices where the observations are binary. In this context, one can discuss the nature or sources of heterogeneity in terms of the following:

1. Response heterogeneity. Different consumers utilize response scales differently. A "4" on a ten point response scale (e.g., preference) may mean different things cognitively to different consumers. Similarly, different consumers may have different predispositions to purchase in the product class. Such types of heterogeneity are typically reflected in the intercept term B_{0it} in equation (1) above.

2. Structural heterogeneity. Different consumers may arrive at similar (or different) levels of utility through different means. That is, individual consumers may value brand attributes differently due to their individual needs. As such, the $\mathbf{B}_{it}$ may vary across consumers reflecting such structural heterogeneity.

3. Perceptual heterogeneity. Consumers may differ in their perceptions, familiarity, and/or recall of the underlying attributes utilized in their decision processes. This can be reflected via different values of the X_{ijkt} akin to an error in variables or stochastic regressors problem.

4. Form heterogeneity. Consumers may also vary in terms of the particular utility function they utilize to evaluate and value the brand attributes. Here, some consumers may use a simple linear model, while others may utilize a more complex non-linear function. Some consumers may process brand attributes in a compensatory fashion, while others in a variety of non-compensatory manners. This particular form of heterogeneity affects F_{it} in equation (1).

5. Distributional heterogeneity. The error term may vary over consumers in one of two major manners. One, the parameters (θ) of the error distribution may be different for different consumers. For example, some groups or segments of consumers may possess higher or lower variance or shape parameters in equation (1). Two, different distributions may be required for different consumers. Thus, such heterogeneity affects the error distribution $g(\theta)$, either with respect to its moments or with respect to the particular distribution itself. (Note that in many choice models there is a direct relationship between the form of $g(\theta)$ and $F(\cdot)$ in equation (1).)

6. Time heterogeneity. Consumers differ in their reactions to their past purchase experiences and behaviors. Some consumers may be more time loyal to a brand than others, where it may take several negative experiences with a brand to invoke switching or a reassessment of attribute importance. Other consumers may possess more volatile utility functions whose structures change over time vis-à-vis word of mouth messages from influential friends or opinion leaders. As such, time heterogeneity can affect virtually any aspect of the utility function posited in equation (1).

1.3. A simple illustration

We illustrate a very simple example of one form of heterogeneity (structural) and the potential costs associated with ignoring it. Let us assume one independent variable or attribute—price, with the dependent variable being probability of purchase. Suppose there are two (unknown) market segments of approximately equal size, and that the utility function is linear in price. For market segment one, members utilize price as a proxy for quality and wish to buy the highest quality good; thus, an upward sloping utility function. Market segment two, on the other hand, can be characterized as extremely price sensitive where the members of this segment wish to purchase the cheapest brand available; thus, a downward sloping utility function. If one were to ignore these fundamental market segment differences and estimate one aggregate utility function, the result is a flat line suggesting that price does not significantly affect purchase probabilities. This aggregate function thus "masks" the truth in this simple example since structural heterogeneity is not considered. (Note that as the number of segments, S, gets large, individual level estimates are obtained in the limit.) The result of neglecting such heterogeneity here is a misrepresentation of the real effects of price on these two separate domains or market segments. (For the marketer, perhaps a sizable profit could be made in solely targeting the quality sensitive segment alone.) While simple plotting routines would have easily illustrated this structure apriori, as the number of attributes and/or segments get large, plotting mechanisms are near impossible to adequately employ.

2. Methodological approaches

2.1. Aggregate level estimation methods

Marketers have benefited from the application of compound distributions from the Bayesian statistical literature (cf. Lilien, Kotler, and Moorthy (1992) for a recent review) to model heterogeneity in consumer responses. Assuming the likelihood $f(y|\mu)$ belongs to the exponential family, and a continuous distribution of some parameter μ: $m(\mu)$, (such as a normal distribution for the mean of a normal, a beta distribution for the probability of the outcome of a binomial, a gamma distribution for the mean of a Poisson, etc.), then a conjugate distribution in the exponential family leads to a form of the compound distribution which is also in the exponential family:

$$h(y|\xi) = \int_{-\infty}^{\infty} f(y|\mu) m(\mu|\xi) d\mu. \tag{2}$$

The Normal-Normal, Beta-Binomial, Negative-Binomial, and Dirichlet-Multinomial have been well known compound distributions applied in Marketing.

In a more formal Bayesian sense, if a conjugate prior distribution for μ, $m(\mu)$, is in the exponential family, then the posterior distribution, $h(\mu|y)$, for that parameter also has the same parametric form as the prior in that exponential family (when $f(y|\mu)$ belongs to the exponential family):

$$h(\mu|y) \propto f(y|\mu)\, m(\mu), \tag{3}$$

where the prior distribution renders some notion of the heterogeneity or the distribution of μ across the population. Such conjugate prior distributions have the practical advantage of computational convenience and are often interpreted as additional information.

If instead of a distribution for the mean, a prior distribution can be formulated for the response coefficients ($\mathbf{B}_{it}$ in (1)); the formulation in equation (2) leads to the class of random coefficient models (cf. Longford 1993), where specified subsets of the coefficients in the response model follow a distribution across the population. For example, suppose:

$$y_i \sim N(\mathbf{X}_i\tilde{\mathbf{B}}_i;\,\theta)$$

and

$$\tilde{\mathbf{B}}_i \sim \mathrm{N}(\mathbf{B}, \underset{\sim}{\Sigma}). \tag{4}$$

Here, the covariance matrix $\underset{\sim}{\Sigma}$ describes between-subjects variation in the **B** coefficients. Those coefficients that are assumed constant correspond to zero variances in the diagonal elements of $\underset{\sim}{\Sigma}$. The corresponding likelihood is:

$$L = \prod_{i=1}^{I} \int_{-\infty}^{\infty} \cdots \int_{-\infty}^{\infty} f(y_i,\tilde{\mathbf{B}}_i)\; m\,(\tilde{\mathbf{B}}_i|\theta) d\tilde{\mathbf{B}}_i, \tag{5}$$

and its maximization is non-trivial for $K > 3$ for non-conjugate $m(\tilde{\mathbf{B}}_i|\theta)$. Zeger and Karim (1991) have applied the Gibbs sampler to this problem to simplify the computational aspects of the estimation. Such an approach can estimate individual level $\mathbf{B}_i$ assuming one has repeated observations per individual (see Allenby and Rossi 1996). An important special case that has been extensively treated in the econometrics literature is where only the intercept term varies randomly in (1).

2.2. Market segment level methods

1. Clusterwise regression. Clusterwise linear regression was a term first coined by Spath (1979) to describe the difficult problem of *simultaneously* finding clusters or market segments (their size and composition), as well as the associated response coefficients by market segment. Consider the following conjoint analysis scenario where:

$\mathbf{X}$	=	$((X_{jk}))$	=	the $(J \times K)$ design matrix of dummy variables;
$\mathbf{Y}$	=	$(\mathbf{y}_1,\mathbf{y}_2,\ldots,\mathbf{y}_I)$	=	the $(J \times I)$ matrix of preference/choice vectors for each of I consumers,
$\mathbf{B}$	=	$(\mathbf{b}_1,\mathbf{b}_2,\ldots,\mathbf{b}_S)$	=	the $(K \times S)$ matrix of conjoint part-worths for the S segments;
$\mathbf{G}$	=	$(\mathbf{g}_1,\mathbf{g}_2,\ldots,\mathbf{g}_I)$	=	the $(S \times I)$ matrix of vectors $\mathbf{G}_i$ defining cluster membership for each consumer i. The individual entries, g_{is}, may be Boolean or fuzzy probabilities, with certain restrictions possible concerning individual membership;
$\mathbf{E}$	=	$(\mathbf{e}_1,\mathbf{e}_2,\ldots,\mathbf{e}_I)$	=	the $(J \times N)$ matrix containing random errors.

The clusterwise regression model for this problem can be formulated as:

$$\mathbf{Y} = \mathbf{XBG} + \mathbf{E}, \tag{6}$$

where given **Y**, **X**, and a value of S (the number of segments), one is interested in estimating **B** and **G** in order to optimize some desirable objective function (e.g., tr(**E**'**E**) as in Spath (1979, 1985)). Hagerty (1985) estimates a transformation or reweighting of the responses to maximize the average expected mean squared error between predictions and validation trials from a hold-out sample. His scheme amounts to a principal axis factor analysis of the correlation matrix of respondents (whose elements are rarely bounded between zero and one). Kamakura (1988) proposed a hierarchical clustering least-squares based procedure to maximize Hagerty's predictive accuracy index where the allocation of subjects to segments is found using an agglomerative clustering procedure. Wedel and Kistemaker (1989) introduced an extension of Spath's (1979) procedure using an exchange algorithm to arrive at a non-overlapping partition of consumers to segments. DeSarbo, Oliver, and Rangaswamy (1989) develop a flexible clusterwise linear regression procedure utilizing simulated annealing discrete optimization algorithms to jointly estimate **B** and **G** in (6). Their methodology can accommodate multiple dependent variables, overlapping or non-overlapping clusters/segments, as well as constraints on cluster membership (cf., DeSarbo and Mahajan, 1984). Wedel and Steenkamp (1989, 1991) propose fuzzy clusterwise linear regression procedures which allow for consumers partial membership in more than one segment, as well as the joint fuzzy clustering of consumers and brands. Krieger and Green (1996) present a K-Means based approach to clusterwise regression to enhance the interpretation of the resulting segments. Finally, DeSarbo and Grissaffe (1996) recently introduce a general clusterwise regression procedure that accommodates multiple objective functions, as well as highly non-linear response functions, utilizing genetic algorithms. See Vriens, Wedel, and Wilms (1996) for a Monte Carlo evaluation of several of these deterministic methods.

2. Latent structure regression. In finite mixture models, it is assumed that a sample of observations arises from a (initially specified) number of underlying classes of unknown proportions. The form of the density of the observations in each of the underlying classes is specified, and the purpose of the finite mixture approach is to decompose the sample into its mixture components. Thus, unlike deterministic clusterwise regression formulations, such finite mixture problems involve parametric distributions. In latent structure regression formulations, one deals with finite mixtures of conditional distributions given covariates **X**. Specifically, we assume a set of multivariate observations on a set of n objects $\mathbf{y}_1,\ldots,\mathbf{y}_I$ as realized values of i.i.d. random variables **Y**. Each $\mathbf{y}_i = (y_{ij})$, ($i = 1,\ldots, I$; $j = 1,\ldots, J$) is a vector of dimension J ($J = 1$ handles the univariate case), which is assumed to arise from a superpopulation which is a mixture of a finite number (S) of segments or classes, G_S ($s = 1,\ldots, S$), in proportions $\pi_1,\ldots, \pi_S$, where it is not known in advance from which class a particular observation arises. The proportions (prior probabilities or mixture weights), π_s, satisfy the following constraints:

$$\sum_{s=1}^{S} \pi_s = 1, \pi_s > 0, s = 1,\ldots,S. \tag{7}$$

The conditional probability density function of $\mathbf{y}_i$ (or conditional mass function in the case of a discrete sample space), given that $\mathbf{y}_i$ comes from class s, is:

$$\mathbf{y}_i \sim f_{i|s}(\mathbf{y}_i;\mathbf{X}_i,\mathbf{B}_{i\in s}). \tag{8}$$

These conditional densities are usually assumed to belong to the same parametric family, although this restriction is not strictly required. The unconditional density of observation i is given by:

$$f_i(\mathbf{y}_i;\phi) = \sum_{s=1}^{S} \pi_s f_{i|s}(\mathbf{y}_i;\mathbf{X}_i,\mathbf{B}_s), \tag{9}$$

where $\phi = (\pi,\Theta)$ denotes the vector of all unknown parameters to be estimated, and $\Theta = (\mathbf{B}_1,\mathbf{B}_2,\ldots, \mathbf{B}_S)$. The random vector $\mathbf{y}_i$ is said to have a finite mixture distribution, with component densities $f_{i|s}(\mathbf{y}_i;\mathbf{X}_i, \mathbf{B}_s)$ and mixing weights π_s. Note that, conditional on sample estimates of π and Θ, the posterior probability of observation i into class s is:

$$P_{is} = \frac{\hat{\pi}_s f_{i|s}(\mathbf{y}_i;\mathbf{X}_i,\hat{\mathbf{B}}_s)}{\sum_{s=1}^{S} \hat{\pi}_s f_{i|s}(\mathbf{y}_i;\mathbf{Y}_i,\hat{\mathbf{B}}_s)}. \tag{10}$$

These P_{is}'s provide a "fuzzy clustering" of the observations into the S segments or classes, and have often been used to classify a given sample of individual observations into groups. Estimation of ϕ is performed via maximum likelihood typically using the *E-M* formulation (cf. Dempster, Laird, and Rubin, 1977). Note, subsets of the parameters may be allowed to vary (or remain fixed) across these mixture components. If only the intercept terms are allowed to vary across mixture components, the traditional econometric approach by Heckman and Singer (1984) is obtained as a special case.

To derive the *E-M* algorithm for such conditional mixture regression models, non-observed data, z_{is}, are introduced, indicating if observation i belongs to segment s: $z_{is} = 1$ if i comes from segment s and $z_{is} = 0$ otherwise. It is assumed that the z_{is} are i.i.d. multinomial, consisting of one draw from the S classes with probabilities $\pi_1,\ldots,\pi_S$. With z_{is} considered as missing data, and assuming that $\mathbf{y}_1,\ldots, \mathbf{y}_I$ given $\mathbf{z}_1,\ldots,\mathbf{z}_I$ are conditionally independent, the complete log-likelihood function can be formed (Dempster, Laird, and Rubin, 1977):

$$\ln L_c(\phi) = \sum_{i=1}^{I} \sum_{s=1}^{S} z_{is} \ln f_{i|s}(\mathbf{y}_i;\mathbf{X}_i,\mathbf{B}_s) + \sum_{i=1}^{I} \sum_{s=1}^{S} z_{is} \ln \pi_s. \tag{11}$$

Using some initial estimate of ϕ, ϕ^0, in the E-step the expectation of ln $L_c(\phi^0)$ is calculated with respect to the conditional distribution of the non-observed data z_{is}, given the observed data $\mathbf{y}_i$ and the provisional estimates ϕ^0. This expectation is obtained by replacing z_{is} in equation (11) by its current expected value, $E\{z_{is}|\mathbf{y},\phi^0\}$, which, using Bayes' rule, is identical to the posterior probability that $\mathbf{y}_i$ belongs to segment s defined in equation (10).

In order to maximize $E\{\ln L_c(\phi)\}$ with respect to ϕ in the M-step, the non-observed data z_{is} in (11) are replaced by their current expectations P_{is}. Maximizing $E\{\ln L_c(\phi)\}$ with respect to π_s, subject to the constraints (in equation 7) on these parameters, yields:

$$\pi_s = \sum_{i=1}^{I} P_{is}/n. \tag{12}$$

Maximizing $E\{\ln L_c(\phi)\}$ with respect to $\boldsymbol{B_s}$ leads to independently solving each of the S expressions:

$$\sum_{i=1}^{I} P_{is}\,(\partial\, ln\, f_{i|s}(\mathbf{y}_i;\, \mathbf{X}_i,\, \mathbf{B}_s)/\partial\mathbf{B}_s) = 0,\ (s = 1, \ldots S). \tag{13}$$

The E- and **M**-steps are alternated until no further improvement in the likelihood function is possible. An attractive feature of the E-M algorithm is that the solution to the M-step often provides closed form expressions for the parameter estimates, such as in the case of the normal density. (Titterington, Smith, and Makov (1985) discuss the general form of the stationary equations for the mixture of distributions from the exponential family.) A second attractive feature of the algorithm is that it provides monotone increasing values of the likelihood (Dempster, Laird and Rubin, 1977). Under mild conditions, the likelihood is bounded from above, so that convergence to at least a local optimum can be established using Jensen's inequality (c.f., Titterington, Smith, and Makov, 1985). Boyles (1983) and Wu (1983) discuss the convergence properties of the E-M algorithm. The problem of multiple maxima of the likelihood of mixture models has been well documented (Titterington, Smith, and Makov, 1985). This difficulty can be dealt with by performing several parameter estimations with different sets of starting values. Wedel and DeSarbo (1994) review the identification requirements in such latent structure regression models, as well as various heuristics for selecting the number of segments or classes. These authors also present an updated list of developments and applications of various forms of latent structure regression.

3. Hybrid approaches

1. Compound finite mixture regression models. We assume S segments with sizes $\pi_1, \ldots, \pi_s$, the random variable y_i distributed as $f_{i|s}(y_i|\mu_s)$, a conjugate distribution $m(\mu_s)$ leading to compound distribution $h_{i|s}(y|\mu_s)$ with expectation μ_s in segment s, and a link function $l(\mu_s) = \eta_s$ with linear predictor $\eta_s = \mathbf{XB_s}$. The likelihood

$$L = \prod_{i=1}^{I} \sum_{s=1}^{S} \pi_s h_{i|s}(y_i|\mathbf{B}_s) \tag{14}$$

is maximized with respect to the segment level $\mathbf{B}_s$. Ramaswamy, Anderson, and DeSarbo (1994) and Bockenholt (1993) have developed such finite mixture compound regression models.

2. Random coefficient finite regression models. This formulation follows a similar mathematical development as with compound finite mixture models concerning segments, distribution, link functions, and linear predictors. However, now one specifies a distribution of the $\mathbf{B}_s$, say $\Phi(\mathbf{B}_s|\theta_S) = N(\mathbf{B}_s^*, \Sigma_s^*)$, where the likelihood becomes:

$$L = \prod_{i=1}^{I} \sum_{s=1}^{S} \pi_s \int_{-\infty}^{\infty} \ldots \int_{-\infty}^{\infty} f_{i|s}(y_i|\mathbf{B}_s)\Phi(\mathbf{B}_s|\theta_S)d\mathbf{B}_s. \tag{15}$$

Lenk and DeSarbo (1996) and Allenby and Ginter (1996) have independently formulated hierarchical Bayes models for a variety of different models. The Gibbs sampler is used to simulate the integrals in (15) since they cannot be analytically evaluated or easily numerically approximated (see Gilks, Richardson, and Spiegelhalter, 1996 for computational details for such Markov Chain Monte Carlo procedures).

In the context of discrete choice models, Gonul and Srinivasan (1993a) incorporate random heterogeneity in intrinsic brand preferences as well as response parameters. Their results underscore the importance of accounting for heterogeneity in the nuisance and the structural parameters of their model. Incorporating state dependence or loyalty is always problematic in the context of random coefficient models. Gonul and Srinivasan (1993a) suggest the mover-stayer method to control for "hard-core" loyal households in the random coefficient model.

A parametric representation of heterogeneity may not truly reflect the underlying distribution of the heterogeneity structure. Alternatively, a semi-parametric method of approximating the underlying distribution through a number of support points (Heckman and Singer 1984; Chintagunta, Jain, and Vilcassim 1991) can be adopted. However, the researcher must be aware that such models may encounter slow convergence problems. More importantly, they may fail to represent the true structure (Heckman and Singer 1984). In a competing-risks hazard model, Gonul and Srinivasan (1993b) offer a direct comparison of the fixed effects with the random effects specification. They find that the Gamma distribution adequately captures the underlying distribution as given by the histogram of the fixed effects estimates.

4. Future research directions

Given the extensive amount of research performed in this exciting area, there exists a number of avenues for future productive research effort:

1. Selection of the number of market segment (S^*). There are no existing statistical tests available in segment level models for selecting S^*—the appropriate number of groups or segments. Heuristics such as AIC, MAIC, BIC, CAIC, ICOMP, NEC, etc. are useful guides, but often suggest different values of S^*. Simulation results by Jedidi, Jagpal, and DeSarbo (1996) and Rust, Simester, Broadie, and Nilikant (1995) suggest that BIC and CAIC are superior heuristics in recovering the true number of segments.

2. Predictive validation in mixture models. It is often difficult to generate accurate predictions to a hold out sample in the mixture model framework, especially in those empirical applications where substantial within segment heterogeneity exists. Because the predictions will, by definition, be a convex combination of the segment centroids, the variance of the predictions is restricted.

3. Model selection. There are few guidelines available as to when finite mixtures vs. compound distributions, for example, are to be preferred. Most of the time, it reduces to "try both and evaluate which one works best". For example, are finite mixture approaches uncovering underlying segments or are they merely approximating (semi-parametrically) a continuous distribution? Extensive Monte Carlo testing is required to develop such guidelines. Preliminary simulation results by Jedidi, Jagpal, and DeSarbo (1996) and Vriens, Wedel, and Wilms (1996) suggest that the finite mixture approach is superior to hierarchical, non-hierarchical, and clusterwise regression clustering in representing more continuous forms of heterogeneity.

4. Profiling segments in mixtures. One of the profound difficulties encountered with finite mixture models is the fact that membership in the derived groups or segments typically relate weakly to any individual demographic or psychographic data. MacReady and Dayton (1980) proposed a reparameterization of the mixing proportions:

$$\pi_{is^*} = \frac{\exp(\mathbf{Z}_i \gamma_{s^*})}{\sum_{s=1}^{S} \exp(\mathbf{Z}_i \gamma_s)}, \tag{16}$$

also utilized in Gupta and Chintagunta (1992) and Kamakura, Wedel, and Agrawal (1994). However, in most applications, the likelihood tends to dominate the prior or mixing distribution, and so the posterior probabilities tend to still be unrelated to these $\mathbf{Z}$ background variables. DeSarbo and Grisaffe (1996), and Krieger and Green (1996) have presented deterministic clusterwise regression alternatives to these mixture approaches that explicitly attempt to take into account such background variables in the objective functions being optimized. Multilevel models allow both demographic and causal variables to be incorporated in such models (see Ansari, Gupta, and Morrin, 1996).

5. Other forms of heterogeneity. The models discussed thus far accommodate response, structural, and limited forms of error heterogeneity. Little has been done to explore issues in form, time, and more complex aspects of error heterogeneity. Kamakura, Kim, and Lee (1996) develop one of the first mixture models to represent different consumer decision processes. The dynamic aspects of changes in market segments was first explored by Poulson (1982), and later by Ramaswamy, Cohen, and Chatterjee (1996). More work is required regarding these more complex forms of heterogeneity, as well as a set of diagnostics to reveal which forms may be present in a particular data set.

6. Endogeneity and simultaneity problems. Jedidi, Jagpal, and DeSarbo (1996) and Jedidi, Ramaswamy, DeSarbo, and Wedel (1996) have recently proposed finite mixture model specifications for general causal model frameworks. Their methodology accommodates recursive models, simultaneous equations, standard and second order factor analysis, and structural equations in a finite mixture setting. Future research should aim at the construction of models to account for consumer heterogeneity involving factors that may be both endogenous and exogenous.

References

Allenby, G. M., and J. Ginter. (1996). On the identification of market segments, *Working Paper*, Ohio State University, Columbus, Ohio.

Allenby, G. M., and P. E. Rossi. (1996). Marketing models of consumer heterogeneity. *Working Paper*, Ohio State University, Columbus, Ohio.

Ansari, A., A. Gupta, and M. Morrin. (1996). Bayesian approaches for modeling multilevel data in management. Working Paper, Columbia University.

Bockenholt, U. (1993). A latent class regression approach for the analysis of recurrent choice data, *British Journal of Mathematical and Statistical Psychology* 46, 95–118.

Boyles, R. A. (1983). On the convergence of the E-M algorithm. *Journal of the Royal Statistical Society B* 45, 47–50.

Chintagunta, P., D. Jain, and N. Vilcassim. (1991). Investigating heterogeneity in brand preferences in logit model for panel data, *Journal of Marketing Research* 28, 417–428.

DeSarbo, W. S., and D. Grisaffe. (1996). Combinatorial optimization approaches to normative market segmentation, *1996 ARTS Forum Presentation*, June 1996.

DeSarbo, W. S., and V. Mahajan. (1984). Constrained classification: the use of a priori information in cluster analysis, *Psychometrika* 49, 187–215.

DeSarbo, W. S., R. Oliver, and A. Rangaswamy. (1989). A simulated annealing methodology for clusterwise linear regression, *Psychometrika* 54(4), 707–736.

DeSarbo, W. S., V. Ramaswamy, and S. Cohen. (1995). Marketing segmentation with choice-based conjoint analysis, *Marketing Letters* 6, 137–148.

Dempster, A. P., N. M. Laird, and R. B. Rubin. (1977). Maximum likelihood from incomplete data via the EM-algorithm. *Journal of the Royal Statistical Society B* 39, 1–38.

Gilks, W. R., S. Richardson, and D. J. Spiegelhalter. (1996). *Markov Chain Monte Carlo in Practice*. London: Chapman and Hall.

Gonul, F., and K. Srinivasan. (1993a). Modeling multiple sources of heterogeneity in multinomial logit models: Methodological and managerial issues, *Marketing Science* 12, 213–229.

Gonul, F., and K. Srinivasan. (1993b). Consumer purchase behavior in a frequently bought product category: Estimation issues and managerial insights from a hazard function model with heterogeneity, *Journal of the American Statistical Association* 88, 1219–1227.

Gupta, S., and P. K. Chintagunta. (1992). On using demographic variables to determine segment membership in logit mixture models, *Journal of Marketing Research* 31, 128–136.

Hagerty, M. R. (1985). Improving the predictive power of conjoint analysis; the use of factor analysis and cluster analysis, *Journal of Marketing Research* 22, 168–184.

Heckman, J. J., and B. Singer. (1984). A method for minimizing the impact of distributional assumptions in econometric models for duration data, *Econometrica* 52 271–320.

Jedidi, K., S. Jagpal, and W. S. DeSarbo. (1996). STEMM: A general finite mixture structural equation model, *Journal of Classification* (forthcoming).

Jedidi, K., H. Jagpal, and W. S. DeSarbo. (1996). Semi-Parametric Structural Equation Models for Response-Based Segmentation and Unobserved Heterogeneity, *Marketing Science* (forthcoming).

Jedidi, K., V. Ramaswamy, W. S. DeSarbo, and M. Wedel. (1996). On estimating finite mixtures of multivariate regression and simultaneous equation models, *Structural Equation Modeling* 3(3), 266–287.

Kamakura, W. A. (1988). A least-squares procedure for benefit segmentation with conjoint experiments, *Journal of Marketing Research* 25, 157–167.

Kamakura, W., B. D. Kim, and J. Lee. (1996). Modeling preference and structural heterogeneity in consumer choice, *Marketing Science* 15-2, 152–172.

Kamakura, W. A., M. Wedel, and J. Agrawal. (1992). Concomitant variable latent class models for the external analysis of choice data, *International Journal of Research in Marketing* 11, 451–464.

Krieger, A., and P. E. Green. (1996). Modifying cluster-based segments to enhance agreement with an exogenous response variable, *Journal of Marketing Research* 33, 351–363.

Lenk, P. J., and W. S. DeSarbo. (1996). Bayesian inference of finite mixture, random effects, generalized linear models, Working Paper, University of Michigan.

Lenk, P. J., W. S. DeSarbo, P. E. Green, and M. R. Young. (1996). Hierarchical Bayes conjoint analysis: recovery of partworth heterogeneity from reduced experimental designs, *Marketing Science* 15-2, 173–191.

Lilien, G. L., P. Kotler, and K. S. Moorthy. (1992). *Marketing Models*. New Jersey: Prentice-Hall Inc.

Longford, N. T. (1993). *Random Coefficient Models*. New York: Oxford University Press.

MacReady, G. B., and C. M. Dayton. (1980). The nature and use of state mastery models. *Applied Psychological Measurement* 4, 493–516.

Poulsen, C. A. (1982). *Latent Structure Analysis with Choice Modeling Applications*. Aarhus: Aarhus School of Business Administration and Economics.

Ramaswamy, V., E. W. Anderson, and W. S. DeSarbo. (1994). Clusterwise negative binomial regression for count data analysis. *Management Science* 40, 405–412.

Ramaswamy, V., R. Chatterjee, and S. Cohen. (1996). Joint segmentation on distinct interdependent bases with categorical data, *Journal of Marketing Research* 33, 337–350.

Rust, R. T., D. Simester, R. J. Broadie, and V. Nilikant. (1995). Model Selection Criteria: An Investigation of Relative Accuracy, Posterior Probabilities, and Combination of Criteria, *Management Science* 41, 322–333.

Späth, H. (1979). Algorithm 39: Clusterwise linear regression. *Computing* 22, 367–373.

Spath, H. (1985). *Cluster Dissection and Analysis Theory*, *FORTRAN Programs*, *Examples*. England: Ellis Horwood Limited.

Titterington, D. M., A. F. M. Smith, and U. E. Makov. (1985). *Statistical Analysis of Finite Mixture Distributions*. New York: Wiley.

Vriens, M., M. Wedel, and T. Wilms. (1996). Metric conjoint segmentation methods: a Monte Carlo comparison, *Journal of Marketing Research* 33, 73–85.

Wedel, M., and W. S. DeSarbo. (1994). A review of latent class regression models and their applications. In Richard P. Bagozzi (ed.), *Advanced Methods for Marketing Research* 353–388.

Wedel, M., and W. S. DeSarbo. (1995). A mixture likelihood approach for generalized linear models, *Journal of Classification* 12, 21–55.

Wedel, M., and C. Kistemaker. (1989). Consumer benefit segmentation using clusterwise linear regression. *International Journal of Research in Marketing* 6, 45–59.

Wedel, M., and J. E. M. Steenkamp. (1989). A fuzzy clusterwise regression approach to benefit segmentation, *International Journal of Research in Marketing* 6(4), 241–258.

Wedel, M., and J. E. M. Steenkamp. (1991). A clusterwise regression method for simultaneous fuzzy market structures and benefit segmentation, *Journal of Marketing Research* 28, 385–396.

Wu, C. F. J. (1983). On the convergence properties of the E-M algorithm. *Annals of Statistics* 11, 95–103.

Zeger, S. L., and M. R. Karim. (1991). Generalized linear models with random effect; a Gibbs sampling approach, *Journal of the American Statistical Association* 86(413), 79–86.