

Issues in the Estimation and Application of Latent Structure Models of Choice

WILLIAM R. DILLON
Workshop Chair, Southern Methodist University

ULF BÖCKENHOLT
University of Illinois

MELINDA SMITH DE BORRERO
Lehigh University

HAM BOZDOGAN
University of Tennessee

WAYNE DE SARBO
University of Michigan

SUNIL GUPTA
Columbia University

WAGNER KAMAKURA
University of Pittsburg

AJITH KUMAR
Arizona State University

BENKATRAM RAMASWAMY
University of Michigan

MICHAEL ZENOR
University of Texas

Key words: latent structure models, discrete choice, finite mixture models

Abstract

Our paper provides a brief review and summary of issues and advances in the use of latent structure and other finite mixture models in the analysis of choice data. Focus is directed to three primary areas: (1) estimation and computational issues, (2) specification and interpretation issues, and (3) future research issues. We comment on what latent structure models have promised, what has been, to date, delivered, and what we should look forward to in the future.

1. Background

Though originally motivated in terms of latent variable models, latent structure models fall under the umbrella of what are referred to as *finite mixture models* (cf.

McLachlan and Basford, 1988; Titterton, Smith, and Makov, 1985). Latent structure analysis is a special case of finite mixture models in which the axiom of local independence is explicitly invoked. The term “finite” is used because the latent continuum is assumed to be finite and discrete. The term “mixture” is used because the population is assumed to consist of homogeneous subgroups. These subgroups are frequently referred to as the *components of the mixture*, or simply as *components*. Let $f(\mathbf{y}, \varphi_s)$ be the probability density for observation \mathbf{Y} , when sampled from the s th component. Assume further that the s th component is a fraction θ_s of the total population, with $\sum_{s=1}^S \theta_s = 1$. If samples from the entire population are taken without knowledge of the component membership of each observation then \mathbf{Y} has the mixture density

$$f(\mathbf{y}, \boldsymbol{\varphi}) = \sum_{s=1}^S \theta_s f(\mathbf{y}, \varphi_s) \quad (1)$$

where the unknown parameter vector $\boldsymbol{\varphi}$ consists of S component parameters $\varphi_1, \dots, \varphi_S$ and S component proportions, often referred to as *mixing proportions* or *mixing weights*, $\theta_1, \dots, \theta_S$. Depending on the form of the density function that is assumed to characterize each component of the mixture, different latent structure models can be realized.

Recently, latent structure and other finite mixture models have proven to be an extremely versatile class of models in the analysis of (scanner) panel, survey, and experimentally derived choice data. The allure and potential promise of latent structure choice models are their ability to account for taste differences without the need for estimation at the individual level. Under a latent structure model individuals are assumed to belong to one of possibly many (but finite) number of homogeneous latent classes that account for the variation in taste distributions across the population of individuals – a (possibly) unique vector of choice-related parameters is estimated for each latent class. In the context of choice modeling, latent structure models provide a compromise between the approach that treats all individuals as different, with a separate set of choice-related parameters estimated for each individual, and the approach that treats all individuals alike, with a single set of choice-related parameters estimated for all individuals.

Several comprehensive reviews of recent applications of latent structure models have appeared. DeSarbo, Manrai, and Manrai (1993) have reviewed latent structure multidimensional scaling models, including latent structure multidimensional scaling models for “pick-any/K” and paired comparison data. DeSarbo and Wedel (1993) have reviewed latent structure regression models, including latent structure regression models for binary choice data and count data. Dillon and Kumar (1992) have reviewed latent class and latent structure models, including recent models appearing in the marketing literature dealing with mixtures of multinomial logits and paired comparison data.

2. Estimation and computational issues

2.1. *ML estimation*

A variety of algorithms can be used to find maximum likelihood solutions. One approach is to apply standard numerical optimization methods to the appropriate set of normal equations that define the within-segment conditional densities. Numerical methods for obtaining ML estimates primarily involve the use of gradient methods. Newton-Raphson, quasi-Newton, simplex, and Fisher's scoring are some of the gradient methods that can be used. Another approach relies on the so-called EM algorithm. In 1977 Dempster, Laird, and Rubin published a seminal paper describing an algorithm that could be used to obtain ML estimates in a variety of situations. The EM algorithm derives its name from the two basic steps in the algorithm. In the E-step (expectation step) new provisional estimates of the component membership probabilities (i.e., the posterior probabilities) are obtained based upon provisional estimates of θ_s and φ_s . In the M-step (maximization step) new estimates of θ_s and φ_s are obtained on the basis of these provisional estimates of the component membership probabilities. These two steps are repeated iteratively.

The primary advantages of numerical optimization procedures are their speed, relative to the EM algorithm, and the ability to easily obtain standard errors for parameter estimates – the usual EM algorithm does not directly provide standard errors. The primary advantages of the EM algorithm are that (1) at each stage in the iterative process the likelihood is monotonically increasing and (2) under certain regularity conditions, the sequence of likelihoods will converge to at least a local maximum. Because of its strong convergence properties we favor use of the EM algorithm.

2.2 *Identifiability*

Problems of identifiability relate to whether unique solutions can be obtained. There are three aspects relating to identifiability that warrant discussion. The first concerns whether a distribution or a family of distributions, parameterized in terms of one or more natural parameters, is generally identifiable. The second concerns the estimation procedure used – whether ML estimation yields identifiable solutions. And the third concerns the particular response model being fit and the class of indeterminacies that are relevant.

The evidence suggests that a large number of distributions are identifiable. Al-Hussaini and Ahmad (1981), for example, have shown that the following families of multivariate distributions yield identifiable mixtures: negative binomial, logarithmic series, Poisson, normal, inverse Gaussian, and random walk. Two exceptions are the general multivariate normal in which the covariances are treated as

unknown parameters (cf. Hathaway, 1983) and mixed-product multinomials (cf. Haberman, 1977, p. 1135).

An arresting feature of latent structure models is that the likelihood function need not be concave. This has two immediate consequences for ML estimation: First, the ML normal equations may fail to maximize the likelihood. Second, multiple maxima of the likelihood may exist: more than one set of ML estimates may maximize the likelihood. The situation is troubling since the EM algorithm guarantees convergence only to a local maximum. This means that alternative starting values may yield strikingly disparate parameter values. Generally, it will be difficult to prove that a particular latent structure model is identifiable. One heuristic is to investigate whether the matrix of second-order partial derivatives of the log-likelihood is negative definite for the parameter estimates obtained at the maximum. If this condition holds, then the solution corresponds to a local maximum of the likelihood.

Because starting values play a role in the evaluation of whether a solution is a global one, how to choose “good” starting values is a question of some concern. The most popular approach to finding starting values is to use the available data. One suggestion is to fit the aggregate model and then use the resulting parameter estimates and associated standard errors to obtain starting values. Another suggestion, implemented in much of the work of DeSarbo (see DeSarbo, Manrai, and Manrai, 1993, and DeSarbo and Wedel, 1993, for relevant references), is to define starting values via quick clustering methods using the available data. In many applications, such methods have proven to render initial solutions quite close to the globally optimum solution. Of course, random starting values can always be used. In situations where lack of convergence may be problematic, the use of random start values may be too ambitious, stretching the limits of the algorithm beyond what is reasonable to expect.

In many latent structure models the natural parameters that define the distributional and model form are themselves reparameterized with respect to a set of covariates. Depending on the form of the reparameterization, identifiability may or may not hold. Proper restrictions must be imposed on parameters for identifiability to hold for each latent segment. In certain instances, latent structure models can help when fitting response models that have known indeterminacies – e.g. unfolding models – by providing a more parsimonious description of the data that thereby allows both internal and external unfolding models to be fit (cf. Böckenholt and Böckenholt, 1991).

2.3 Determining the number of latent classes/segments

The overwhelming temptation in deciding how many components to retain is to compute the likelihood ratio statistic between models having S and $S + 1$ components. Unfortunately, in the case of mixtures the likelihood ratio statistic is not

asymptotically χ^2 -distributed with known degrees of freedom given by the difference in the number of parameters estimated under the null and alternative hypotheses. Because of the limitations of the likelihood ratio statistic other alternatives have been suggested. The following five alternatives have gained some popularity.

1. One alternative is to use parametric bootstrapping (Hope, 1968). The procedure involves drawing $T - 1$ random Monte Carlo samples of size N from a population having S -components and prespecified density, and fitting S and $S + 1$ component models to each of the generated samples. Anyone that has used this method knows all too well that it is extremely costly in terms of machine time.
2. A second alternative is to use *information-theoretic*-based model selection criteria to assess goodness-of-fit. With this approach the decision on the number of components to retain proceeds by minimizing a particular information-theoretic criterion. The oldest of these information criteria measures is AIC (Akaike, 1974). Other criteria include BIC (Schwarz, 1978) and CAIC (Bozdogan, 1987), which attempt to penalize overparameterization more severely than does AIC. Though the AIC, BIC, and CAIC have a long history of use, recent simulation studies suggest that they tend to systematically underfit or overfit depending on sample size and component separation (cf. de Borrero, 1993).
3. Bozdogan (1993) developed a new entropic complexity criterion called ICOMP. Similar to AIC, BIC, and CAIC, ICOMP is a penalized information-based validity functional. It can be viewed in terms of two distinct components. The first component measures the lack of fit of the model, and the second component measures the complexity of the estimated inverse-Fisher information matrix. The preliminary evidence on the performance of the ICOMP measure is quite encouraging. More comprehensive and intense simulations are, however, needed before definitive and generalizable conclusions can be drawn.
4. The fourth alternative, introduced by Windham and Cutler (1992), is called the *minimum information ratio* (MIR). The MIR is the smallest eigenvalue of the matrix $F_c^{-1} F$, where F is the Fisher information matrix, and F_c is the Fisher information matrix for the classified sample. The MIR can be viewed as a measure of the proportion of information about the parameters available without knowledge of the subpopulation membership of the sample. The MIR may not at all perform well: see de Borrero (1993) and Cutler and Windham (1993).
5. The fifth alternative represents a very different approach. It involves what is called nonparametric ML estimation (NPMLE) (cf. Böhning, Schlattmann, and Lindsay, 1992), or what is also called the *flexible support size* method since S itself is treated as an unknown parameter and \mathcal{P} is viewed as completely unknown varying in the set of *all* probability measures. In essence, with NPMLE a grid of parameter values is specified over which the corresponding population proportions that maximizes the likelihood function are found. There are several troubling issues related to the NPMLE approach. First, this ap-

proach is based upon saturating the likelihood with possibly as many points as sample observations, which runs contrary to the basic tenet of latent structure modeling. Second, and perhaps more serious, the algorithms available to date are for univariate distributions only. *Adaptive grid search* procedures are needed in order to extend this approach to multivariate problem settings.

3. Specification and interpretation issues

3.1. Perspectives on heterogeneity

There are quite different perspectives on how consumer heterogeneity can be viewed as the following illustration demonstrates. Let y_{ij} denote the number of selections of brand j by respondent i in some fixed time period, where $y_{ij} \geq 0$, and $\sum_{j=1}^J y_{ij} = y_{i+}$ is the total number of choices by respondent i . Under the assumption that the probability of making a choice is unaffected by the length of time since the last choice was made, the total number of choices y_{i+} can be described by a Poisson process with mean rate λ_i

$$\Pr(y_{i+}; \lambda) = \frac{\exp(-\lambda_i)\lambda_i^{y_{i+}}}{y_{i+}!} \quad y_{i+} \in \{0, 1, \dots\}. \quad (2)$$

Let us further assume the availability of a covariate, denoted by X , which presumably influences the number of choices by each respondent. To account for the effect of a respondent-specific covariate, the rate parameter λ_i can be parameterized as

$$\lambda_i = C_i \exp(X_i\beta), \quad (3)$$

where β is an unknown regression weight and C_i is some measure of the time interval for the i th respondent.

There can be a pronounced lack of fit of the data to this model. Generally, this is a signal that the sample may be too heterogeneous for this specification. In essence, the specification shown in equation 2 may “force” too much structure on the data. There are at least two approaches for accounting for respondent heterogeneity. One approach to accounting for unobserved heterogeneity is to reparameterize λ_i as

$$\lambda_i = C_i \exp(X_i\beta + \gamma_i) = C_i\alpha_i\lambda_i, \quad (4)$$

where γ_i , which can be either random or fixed, captures the heterogeneity in respondent i 's choices. The second approach, consistent with a latent structure model, is to adopt the following reparameterization

$$\lambda_{i|s} = C_i \exp(X_i \beta_s), \quad (5)$$

where $\lambda_{i|s}$ is the mean purchase rate for individual i who belongs to latent segment s and β_s is the value of the unknown regression weight in latent segment s , where $s = 1, 2, \dots, S$.

In the first approach, γ_i is viewed as a nuisance parameter. It provides no meaningful diagnostic information in the sense that it is introduced to capture those factors that have been left out of the model. It is introduced to "overcome" the heterogeneity problem, rather than as a possible explanation. In contrast, the tacit assumption underlying the second approach is that the population consists of a discrete number of homogeneous subgroups or component populations and a simple Poisson model is assumed to hold in each component, where each component of the mixture may have a different parameter (β_s) value. Thus, this perspective views heterogeneity as providing diagnostic information; the latent segments, through the S - β_s weights, capture respondent heterogeneity in response sensitivity.

3.2. Continuous versus discrete latent segments

In general, the unconditional likelihood for a consumer h with choice vector \mathbf{y} can be written as

$$L_h(\mathbf{y}) = \int_{-\infty}^{+\infty} L_h(\mathbf{y} | \varphi) dF(\varphi) \quad (6)$$

where $L_h(\mathbf{y} | \varphi)$ is the conditional likelihood with parameter φ , and $F(\cdot)$ is the mixing distribution. The distinction between discrete and continuous latent constructs may not be particularly important since it can be shown that a continuous mixing distribution function $F(\cdot)$ can be consistently estimated with a finite number of S mass points (cf. Lindsay, 1983).

The assumption that the latent segments account for all the covariation in the data is probably unrealistic: respondents who belong to the same latent segment are to some degree different and thus some within-segment heterogeneity is present. One way to overcome this problem is to incorporate parameters that can account for within-segment variation. For example Böckenholt (1993a) recently presents a latent structure model in which unexplained variations in the choice

probabilities within each latent segment are (potentially) captured by compounding the multinomial distribution with the Dirichlet distribution (Johnson and Kotz, 1969). Under this specification the choice probabilities vary according to the Dirichlet distribution from consumer to consumer within each latent segment and the effects of covariates are modeled by reparameterizing the Dirichlet parameters. Ramaswamy, Anderson, and DeSarbo (1993) present a mixed NBD model that can also account for otherwise unexplained within-segment variation.

4.3. State dependence

When scanner panel data are used to calibrate probabilistic choice models it is generally assumed that successive purchases for a household are independent. Latent structure probabilistic choice models are typically parameterized to be consistent with the proposition that, conditional on latent segment membership, consumers follow a zero-order brand switching pattern. When there is feedback, choice on one occasion affects probabilities of choice on future occasions; some sort of state dependence is operating. State dependence conceivably can manifest several different types of behavior. Brand loyalty, as evinced in large repeat purchase probabilities, for example, could reflect state dependence: the choice of alternative A increases the probability of choosing alternative A on the next occasion. Similarly, variety seeking could reflect state dependence as well: the choice of alternative A decreases the probability of choosing alternative A on the next occasion.

Brand loyalty and brand switching can be distinguished in latent structure choice models. For example, both Grover and Srinivasan (1987) and Dillon and Gupta (1994) advance parametric representations that distinguish between brand loyalty and brand switching. A way of relaxing the zero-order, no feedback assumption is to adopt a mixed Markov chain model. Mixed Markov models come about by conditioning the choice probabilities by the outcomes of the previous choice – i.e., outcomes of the previous choices influence the probability distribution of the next choice. Thus, in a three-wave problem setting we parameterize the probability of purchasing brand *i* on the first occasion, brand *j* on the second occasion, and brand *k* on the third occasion. Mixed Markov models were first introduced by Poulsen (1982). Recently Van de Pol and Langeheine (1989) have made further advances and provide software for fitting this class of models. Note, however, these models do not consider covariates, so they are somewhat limited in the amount of diagnostic information provided.

Another example of a latent structure model with state dependence is the finite-mixture of hazard function models recently suggested by Wedel, Kamakura, and DeSarbo (1993). State dependence is explicitly modeled through a flexible Box-Cox baseline hazard and nonproportional effects of marketing variables, which are estimated through a finite-mixture of Poisson regressions.

4.4. *Nonstationarity*

Stationarity refers to the assumption that consumers do not change their latent segment membership over time. The assumption of stationarity seems overly restrictive in at least two ways. First, if consumers are inextricably fixed to a latent segment, then segments are, by definition, presumed not to grow or diminish over time. Second, this assumption implies that underlying preference structures must be constant and thus are not influenced by marketing mix activity (advertising, promotion), changes in demographics or environmental factors (aging population, weather), politics (changing tax code), etc.

The only nonstationary latent structure probabilistic choice models that have appeared are due to the work of Poulsen (1982) and extensions by Van de Pol and Langeheine (1989). These models fall under the umbrella of latent Markov models. The driving premise underlying these models is that the observed choices are imperfect indicators, reflecting some measurement or uncertainty, and that changes are indeed taking place which cause shifts between the latent states. Thus, at any purchase occasion each consumer is in a particular state; however, over a number of purchase occasions, consumers are allowed to change their latent state position. The movement between latent states is governed by a stationary transition matrix that is conditioned on the present state the consumer is in. Similar to mixed Markov models, latent Markov models have not, to date, been extended to allow covariates to be included in the parameterization. Finally, a latent structure probabilistic choice model that can account for nonstationarity and state dependence has recently been presented by Böckenholt (1993b).

4.5. *Joint versus concomitant specifications*

In the context of latent structure probabilistic choice models, there have been two approaches to incorporating covariates. One approach includes all covariates in the specification of the within-segment (conditional) density. Thus, a (possibly) unique within-segment coefficient is estimated for each covariate. The other approach, which has been recently suggested (cf. Dayton and MacReady, 1988; Dillon, Kumar, and de Borrero, 1993; Kamakura, Wedel, and Agrawal, 1992), falls under the umbrella of concomitant variable modeling. Concomitant variable models establish a functional linkage between specific covariates and the prior probabilities of latent segment membership – the mixing weights.

The rationale behind concomitant variable modeling is that the covariates available are qualitatively different. For example, in the context of choice modeling, one set of covariates typically includes brand-specific variables (price, brand attribute ratings), and another set might pertain to respondent-based variables (income, family size). The former set of covariates may be thought to influence choice directly in the sense that they establish the defining character of the latent segment (a price-sensitive segment, a health-conscious segment, etc.), whereas

the latter set conceivably can be viewed as concomitant variables since they are conceptualized to influence the latent segment membership of a respondent, but are not brand-specific.

5. Future research agenda

We hope the reader will come away with an appreciation of what latent structure probabilistic models have to offer as well as with heightened sensibilities concerning future research issues. Thus, a fitting close to our discussion is to focus on future research priorities:

1. Develop adaptive grid search procedures to implement nonparametric maximum likelihood estimation with multivariate distributions.
2. Identify estimation algorithms that provide high statistical efficiency and low search.
3. Provide comprehensive simulations comparing various methods for determining the number of latent components to retain with particular focus on ICOMP and different methods for computing the Fisher information matrix.
4. Initiate comparative simulation studies focusing on alternative methods for generating starting values.
5. Develop latent structure models that can account for consumer heterogeneity caused by hierarchical data structures.
6. Understand the estimation costs and diagnostic benefits of explicitly specifying latent loyalty segments.
7. Develop latent structure models that can incorporate prior information in a Bayesian framework.
8. Provide more flexible and comprehensive latent structure probabilistic choice models in which consumers are allowed to change latent segment membership over time as a function of covariates.
9. Develop latent structure probabilistic choice models that relax the zero-order assumption and parameterize state dependence in terms of a set of covariates.

References

- Akaike, H. (1974). "A New Look at Statistical Model Identification," *IEEE Transactions on Automatic Control*, AC-19, 716–723.
- Al-Hussani, E. K., and K. E. Ahmad. (1981). "On the Identifiability of Finite Mixtures of Distributions," *IEEE Transactions on Information Theory*, 27, 664–668.
- Böckenholt, U. (1993a). "Estimating Latent Distributions in Recurrent Choice Data," *Psychometrika*, 58, 489–509.
- . (1993b). "Latent Change in Recurrent Choice Data," Unpublished manuscript, Department of Psychology, University of Illinois.

- Böckenholt, U., and I. Bockenholt. (1991). "Constrained Latent Class Analysis: Simultaneous Classification and Scaling of Discrete Choice Data," *Psychometrika*, 56, 699-716.
- Böhning, D., P. Schlattmann, and B. G. Lindsay. (1992). "Computer Assisted Analysis of Mixtures (C.A.MAN): Statistical Algorithms," *Biometrics*, 48, 283-303.
- Bozdogan, Hamparsum. (1987). "Model Selection and Akaike's Information Criterion (AIC): The General Theory and Its Analytical Extensions," *Psychometrika*, 52, 345-370.
- . (1993). "Mixture-Model Cluster Analysis Using Model Selection Criteria and a New Informational Measure of Complexity." In H. Bozdogan (ed.), *Proceedings of the First U.S./Japan Conference on the Frontiers of Statistical Modeling, Vol. 2: Multivariate Statistical Modeling*. Dordrecht, the Netherlands: Kluwer Academic Publishers.
- Cutler, Adele, and Michael P. Windham. (1993). "Information-Based Validity Functionals for Mixture Analysis." In H. Bozdogan (ed.), *Proceedings of the First U.S. Japan Conference on the Frontiers of Statistical Modeling, Vol. 2: Multivariate Statistical Modeling*. Dordrecht, the Netherlands: Kluwer Academic Publishers.
- Dayton, C. Mitchell, and George B. MacReady. (1988). "Concomitant-Variable Latent-Class Models," *Journal of the American Statistical Association*, vol. 83, no. 401, Theory and Methods.
- de Borrero, Melinda. (1993). "Determining the Number of Mixture Components and a Poisson-Multinomial Logit Mixture Model Application," Unpublished dissertation, University of South Carolina.
- Dempster, A. P., N. M. Laird, and D. B. Rubin. (1977). "Maximum Likelihood from Incomplete Data via the EM Algorithm," *Journal of Royal Statistical Society*, B39, 1-38.
- DeSarbo, Wayne S., Ajay K. Manrai, and Lalita A. Manrai. (1993). "Latent Class Multidimensional Scaling: A Review of Recent Developments in the Marketing and Psychometric Literature," In R. Bagozzi (ed.), *The Handbook of Marketing Research*. London: Blackwell Publishers.
- DeSarbo, Wayne S., and Michel Wedel. (1993). "A Review of Recent Developments in Latent Class Regression Models." In Rick P. Bagozzi (ed.), *Handbook of Marketing Research*, forthcoming.
- Dillon, William R., and A. Kumar. (1992). "Latent Structure and Other Mixture Models in Marketing: An Integrative Survey and Overview." In Rick A. Bagozzi (ed.), *Handbook of Marketing Research*, London: Blackwell Publishers.
- Dillon, William R., A. Kumar, and M. Smith de Borrero. (1993). "Capturing Individual Differences in Paired Comparisons: An Extended BTL Model Incorporating Descriptor Variables," *Journal of Marketing Research*, 30, 42-51.
- Dillon, William R., and Sunil Gupta. (1994). "A Segment-Level Model of Category Volume and Brand Choice," Unpublished manuscript, Southern Methodist University.
- Grover, Rajiv, and V. Srinivasan. (1987). "A Simultaneous Approach to Market Segmentation and Market Structuring," *Journal of Marketing Research*, 24, 139-153.
- Haberman, Shelby J. (1977). "Product Models for Frequency Tables Involving Indirect Observation," *Annals of Statistics*, 5(6), 1124-1147.
- Hathaway, R. J. (1983). "Constrained Maximum Likelihood Estimation for a Mixture of Multivariate Normal Densities," Tech. Rep. 92, Department of Mathematics Statistics, University of South Carolina, Columbia, SC.
- Hope, A. C. A. (1968). "A Simplified Monte Carlo Significance Test Procedure," *Journal of the Royal Statistical Society*, Series B, 30, 582-598.
- Johnson, N. L., and S. Kotz. (1969). *Distributions in Statistics: Discrete Distributions*. New York: Wiley.
- Kamakura, Wagner A., Michel Wedel, and Jagdish Agrawal. (1992). "Concomitant Variable Latent Class Models for the External Analysis of Choice Data," Research Memorandum 486, Faculty of Economics, University of Groningen.
- Langeheine, R. (1988). *New Developments in Latent Class Theory, Latent Trait and Latent Class Models*. New York: Plenum.

- Lindsay, B. G. (1983). "The Geometry of Mixture Likelihoods, II: The Exponential Family," *Annals of Statistics*, 11, 783-792.
- McLachlan, G. J., and K. E. Basford. (1988). *Mixture Models: Inference and Application to Clustering*. New York: Marcel Dekker.
- Poulsen, C. A. (1982). "Latent Structure Analysis with Choice Modelling Applications." Aarhus: Aarhus School of Business Administration and Economics.
- Ramaswamy, V., E. W. Anderson, and W. S. DeSarbo. (1993). "A New Latent Class Methodology for Modeling Brand Purchase Frequencies and Deriving Response-Bases Market Segments," *Management Science*, 40, 405-417.
- Schwarz, G. (1978). "Estimating the Dimension of a Model," *Ann Statist*, 6, 461-461.
- Titterton, D. M., A. F. M. Smith, and U. E. Makov. (1985). *Statistical Analysis of Finite Mixture Distributions*. New York: Wiley.
- Van de Pol, F., and R. Langeheine. (1989). "Mixed Markov Latent Class Models," *Sociological Methods and Research*, 15, 118-141.
- Wedel, Michel, Wagner A. Kamakura, and Wayne S. DeSarbo. (1993). "Heterogeneous and Non-proportional Effects of Marketing Variables on Brand Switching: An Application of a Discrete Time Duration Model," Research Note 541, Faculty of Economics, University of Groningen.
- Windham, M. P., and A. Cutler. (1992). "Information Ratios for Validating Mixture Analyses," *Journal of the American Statistical Association*, 87(420), 1188-1192.