10

A Review of Recent Developments in Latent Class Regression Models

Michel Wedel and Wayne S. DeSarbo

Latent Class Mixture Models

The development of mixture models can be historically traced back to the work of Newcomb (1886) and Pearson (1894). Mixture distributions have been of considerable interest in recent years leading to a vast number of methodological and applied papers, as well as to three dedicated monographs (cf. Everitt and Hand, 1981; Titterington, Smith, and Makov, 1985; and McLachlan and Basford, 1988). In finite mixture models, it is assumed that a sample of observations arises from a (initially specified) number of underlying classes of unknown proportions. A concrete 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. Specifically, we assume a set of multivariate observations on a set of n objects $\mathbf{y}_1, \ldots, \mathbf{y}_n$ as realized values of i.i.d. random variables $\mathbf{Y}$. Each $\mathbf{y}_i = ((y_{ij}))$, $(i = 1, \ldots, n; j = 1, \ldots, J)$ 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 groups 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:

$$\sum_{s=1}^{S} \pi_s = 1, \qquad \pi_s > 0, \ s = 1, \ldots, S, \tag{10.1}$$

(the positivity constraints then can be relaxed to allow for negative mixture weights as in Titterington, Smith, and Makov, 1985). 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:

$$\mathbf{y}_i \sim f_{i|s}(\mathbf{y}_i; \boldsymbol{\theta}_s). \tag{10.2}$$

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:

$$f_i(\mathbf{y}_i; \boldsymbol{\varphi}) = \sum_{s=1}^{S} \pi_s f_{i|s}(\mathbf{y}_i; \boldsymbol{\theta}_s), \tag{10.3}$$

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

$$\alpha_{is} = \frac{\hat{\pi}_s f_{i|s}(\mathbf{y}_i; \hat{\boldsymbol{\theta}}_s)}{\sum_{s=1}^{S} \hat{\pi}_s f_{i|s}(\mathbf{y}_i; \hat{\boldsymbol{\theta}}_s)}. \tag{10.4}$$

These α_{is}'s provide a "fluzzy clustering" of the observations into the S groups or classes and have often been used to classify a given sample into groups.

A large variety of parametric forms for the mixture components have been assumed in the literature, including discrete distributions such as the binomial (John, 1970a; Hasselblad, 1966), geometric (Harris, 1983), negative binomial (John, 1970a), hypergeometric (John, 1970a), poisson (John, 1970a; Hasselblad, 1969), and continuous densities such as the (univariate and multivariate) normal (Hasselblad, 1966; Wolfe, 1970), uniform (Gupta and Miyawaki, 1978), exponential (Everitt and Hand, 1981; Teicher, 1961), weibull (Mandelbaum and Harris, 1982), gamma (John, 1970b), and dirichlet (Antoniak, 1974), as well as several compound and truncated distributions. (Croon (1989), and Kamakura and Novak (1992) develop parametric mixture models for rank-ordered data.) A comprehensive overview of applications of specific parametric mixtures is provided by Titterington, Smith, and Makov (1985).

An interesting special case of mixture models is latent structure analysis which is founded upon the concept of local independence (Lazarsfeld and Henry, 1968). In latent structure analysis, the covariance between the components of the multivariate observations y_{ij} are assumed to be caused by the latent classes to which the observations belong. Within latent classes, the component mixtures are assumed to be independent:

$$f_{i|s}(\mathbf{y}_i; \boldsymbol{\theta}_s) = \prod_{j=1}^{J} f_{ij|s}(y_{ij}; \boldsymbol{\theta}_s). \tag{10.5}$$

Lazarsfeld and Henry (1968) provided the first systematic treatment of latent structure analysis. Latent class models are subsumed under the general class of latent structure models and pertain to manifest categorical variables, the associations among which are explained by an unobserved discrete variable (see Dillon's, 1993, review in Chapter 9 for further details). In related latent trait models, associations among manifest categorical variables are explained from an underlying continuous construct (Langeheine, 1988). Both of these procedures are based on the axiom of local independence and have been extensively used in psychological test-theory (Langeheine and Rost, 1988). Goodman (1974a, b) first demonstrated how latent class models could be formulated in a general log-linear model framework, which could be estimated by maximum likelihood. Other contributions to the theory of latent structure analysis have been made by Haberman (1979), Clogg (1979), Goodman (1978, 1979), and Clogg and Goodman (1984, 1986). Textbooks dealing with this topic include those by Everitt (1984), Formann (1984), Bartholomew (1987), and Hagenaars (1990). Some recent developments can be found in the volume edited by Langeheine and Rost (1988).

Identifiability

A parametric family of mixtures is said to be identifiable if distinct values of the parameters determine distinct members of the mixture (McLachlan and Basford, 1988). Titterington, Smith, and Makov (1985) provide an overview of the identifiability of mixtures with specific component densities, demonstrating identifiability of such mixtures involving the normal, poisson, gamma, and the binomial. From a survey of the literature, they conclude that, apart from special cases with finite sample spaces or very special simple density functions (such as mixtures of uniform distributions and S-mixtures of binomial (n, p) distributions with $n < 2S - 1$), identifiability of classes of finite mixtures is generally assured. McLachlan and Basford (1988) have shown that the likelihood is invariant under interchanging of the labels of the latent classes, and propose to report only one of the possible arrangements of the classes, in order to alleviate this threat to identification. Local identification of mixture models can be established from the expected matrix of the second derivatives of the log-likelihood (Formann, 1992). Recent work on the identifiability of mixtures has been reported in Li and Sedransk (1988) and Lüxmann-Ellinghaus (1987).

Estimation

The purpose is to estimate the parameters of the finite mixture, φ, given y_{ij} and a value of S. To accomplish this, several methods can be used. Initially estimation of mixture models proceeded using the method of moments (Pearson, 1894; Charlier and Wicksell, 1924; Quandt and Ramsey, 1978), but later attention focussed on graphical techniques for detection of (univariate) mixtures (see Harding, 1948; Cassie, 1954; Bhattacharya, 1967; Fowlkes, 1979). (An overview of other estimating methods including the Bayesian, minimum distance, and recursive methods is provided by Titterington, Smith, and Makov, 1985). Hasselblad (1966, 1969) was one of the first to use maximum likelihood estimation for mixtures of two or more multivariate normals, as well as other distributions from the exponential family. As maximum likelihood has been found to be generally superior to the method of moments for the estimation of finite mixtures (Fryer and Robertson, 1972), the likelihood approach has become increasingly popular (McLachlan, 1982; Basford and McLachlan, 1985) and appears presently to be the most frequently used method (see Titterington, Smith, and Makov, 1985, for a review of its use). The likelihood for φ can be formulated as:

$$L(\varphi|\mathbf{y}) = \prod_{i=1}^{n} f_i(\mathbf{y}_i; \varphi). \tag{10.6}$$

An estimate of φ can be obtained by maximizing the likelihood equation with respect to φ subject to the restrictions (1). This can be accomplished primarily in three ways: by using standard optimization routines such as the Newton-Raphson or Quasi Newton methods (see McHugh, 1956, 1958), the Method of Scoring (see Titterington, Smith, and Makov, 1985), or the Expectation-Maximization (EM) algorithm (Dempster, Laird, and Rubin, 1977). The Newton-Raphson and scoring methods require relatively few iterations to converge, and provide the asymptotic variances of the parameter estimates as a by-product, but convergence is not ensured. The EM can be programmed easily, iterations are computationally attractive, and convergence is ensured, but the algorithm often requires many iterations (see Titterington, Smith, and Makov, 1985; McLachlan and Basford, 1988). It is as yet unclear which of the two methods is to be preferred in general (see Langenheine, 1988; Mooijaart and van der Heijden, 1992), but the EM algorithm has apparently been some what more popular (Titterington, 1990). While most of the mixture likelihood approaches (beginning with Newcomb, 1886, and later Hasselblad 1966, 1969, and Wolfe 1970) have used iterative schemes corresponding to special cases of the EM algorithm, the formal applicability of this EM algorithm, with its attractive convergence properties, to finite mixture problems was recognized only after the developments of Dempster, Laird, and Rubin (1977), which were later supplemented by Boyles (1983), Wu (1983), and Redner and Walker (1984).

To derive the EM algorithm for latent class regression models, non-observed data, z_{is}, are introduced, indicating if observation i belongs to latent class s: $z_{is} = 1$ if i comes from latent class s and $z_{is} = 0$ otherwise. It is assumed that the z_{is} are i.i.d. multinomial, consisting of one draw on the S classes $G_1, \ldots, G_S$, with probabilities $\pi_1, \ldots, \pi_S$. With z_{is} considered as missing data, and assuming that $\mathbf{y}_1, \ldots, \mathbf{y}_n$ given $\mathbf{z}_1, \ldots, \mathbf{z}_n$ are conditionally independent, the complete log-likelihood function can be formed (Dempster, Laird, and Rubin, 1977):

$$ln\, L_c(\varphi) = \sum_{i=1}^{n} \sum_{s=1}^{S} z_{is}\, ln\, f_{i|s}(\mathbf{y}_i; \boldsymbol{\theta}_s) + \sum_{i=1}^{n} \sum_{s=1}^{S} z_{is}\, ln\, \pi_s. \tag{10.4}$$

This complete log-likelihood is maximized using the iterative EM-algorithm. Using some initial estimate of φ, $\varphi^{(0)}$, in the E-step the expectation of $L_c(\varphi^{(0)})$ 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 $\varphi^{(0)}$. It can easily be seen that this expectation is obtained by replacing z_{is} in equation (10.7) by its current expected value, $E\{z_{is}|\mathbf{y}, \varphi^{(0)}\}$, which, using Bayes' rule, can be shown to be identical to the posterior probability that $\mathbf{y}_i$ belongs to class s defined in equation (10.4).

In order to maximize $E\{\ln L_c(\varphi)\}$ with respect to φ in the M-step, the non-observed data z_{is} in (10.7) are replaced by their current expectations α_{is}. Maximizing $E\{\ln L_c(\varphi)\}$ with respect to π_s, subject to the constraints (in equation (10.1)) on these parameters, yields:

$$\hat{\pi}_s = \sum_{i=1}^{n} \hat{\alpha}_{is}/n. \tag{10.8}$$

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

$$\sum_{i=1}^{n} \hat{\alpha}_{is}\, (\partial\, ln\, f_{i|s}(\mathbf{y}_i; \boldsymbol{\theta}_s)/\partial \boldsymbol{\theta}_s) = 0, \qquad (s = 1, \ldots, S) \tag{10.9}$$

The E-and M-steps are alternated until no further improvement in the likelihood function is possible. Schematically, the EM algorithm may be described as follows:

1 Initialize the iteration index h, $h \leftarrow 0$; read the user specified value of S, and generate a starting partition, $\hat{\alpha}_{is}^{(0)}$. A random starting partition may be obtained or a rational start may be used.
2 Given the $\hat{\alpha}_{is}^{(h)}$, calculate M.L. estimates of the parameters of each component mixtures from equations (10.8) and (10.9)
3 Convergence test: if the change in the log-likelihood from iteration (h) to $(h + 1)$ is smaller than some positive constant, stop.
4 Increment iteration index: $h \leftarrow h + 1$, recalculate the posterior probabilities, $\alpha_{is}^{(h+1)}$, via equation (10.4) and return to step (2).

An attractive feature of the EM-algorithm is that the solution to the M-step (step (2) above) 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 (cf. Titterington, Smith, and Makov, 1985). Boyles (1983) and Wu (1983) provide a discussion of the convergence properties of the EM algorithm. The problem of multiple maxima of the likelihood of mixture models is well documented (Titterington, Smith, and Makov, 1985). This problem can be minimized by performing several parameter estimations with different sets of starting values. Note, Hamilton (1991) proposed a Quasi-Bayesian approach to estimate univariate and multivariate normal mixtures, which consistently improves the maximum likelihood approach for this class of models. It eliminates the singularities associated with maximum likelihood estimation of normal mixtures, while it offers guidance for choosing among locally optimal solutions. Hamilton (1991) provides a Monte Carlo study which supports the performance of the Quasi-Bayesian over the maximum likelihood approach. The estimation equations are solved using an EM algorithm.

The preceding EM algorithm provides a general framework for estimating the various latent class regression models reviewed in the next two sections of this chapter. For these respective models (in step (2) of the algorithm), the derivatives (10.9) are taken with respect to the corresponding regression model parameters specified for each latent class. The specific estimation sub-routine employed in step (2) is dependent upon the structure of these models and the specification of the conditional densities. These issues will be further detailed in the respective sections of this chapter.

Tests for the number of classes

When applying mixture models to empirical data, the actual number of classes, S, is typically unknown. The problem of identifying the number of classes is the inference problem in mixture models with the least satisfactory statistical treatment (Titterington, 1990). Both Titterington, Smith, and Makov (1985) and McLachlan and Basford (1988) each devote an entire chapter to this topic. The problem is that the standard generalized likelihood ratio statistic to test the null-hypothesis (H_0) of S classes against the alternative hypothesis (H_1) of $S+1$ classes is not asymptotically distributed as chisquare, because H_0 corresponds to a boundary of the parameter space for H_1, so that under H_0 the generalized likelihood ratio test statistic is not (asymptotically) a full rank quadratic form (Aitkin and Rubin, 1985; Ghosh and Sen, 1985; Li and Sedransk, 1988; Titterington, 1990).

Various tests for determining the number of classes have been proposed for special types of mixtures (see Titterington, Smith, and Makov, 1985; Anderson, 1985; Henna, 1985; Yarmal-Vuarl and Ataman, 1987). Monte Carlo test procedures have been applied to mixture problems by Aitkin, Anderson, and Hinde (1981), and McLachlan (1987). These procedures involves comparing the likelihood ratio statistic for $S + 1$ versus S classes from the real data with a distribution of that statistic obtained from K datasets containing S classes, which are generated by replacing the unknown parameters in the densities by their likelihood estimates from the original data. This procedure is computationally very cumbersome however (cf. McLachlan and Basford, 1988), while the observed rejection rates do not quite conform to the intended levels under the null-hypothesis (Titterington, 1990).

Another class of techniques for testing the number of components present are those based on information criteria in which a penalty, proportional to the number of parameters estimated, is imposed on the maximized log-likelihood:

$$C = -2 \ln L - t\mathrm{d}, \tag{10.10}$$

where t denotes the number of parameters estimated, and d is some constant. Sclove (1977) and Bozdogan and Sclove (1984) proposed the use of Akaike's Information criterion (AIC; Akaike, 1974) to determine the number of classes; with the AIC, $\mathrm{d} = 2$ in equation (10.10). Two criteria related to the AIC statistic are Schwartz's (1978) Bayesian Information Criterion (BIC), where $\mathrm{d} = \ln(n)$, and Bozdogan's (1987) consistent AIC (CAIC), where $\mathrm{d} = \ln(n) + 1$. Both of these statistics impose an additional sample size penalty on the log-likelihood, and are more conservative than the AIC statistic tending to favor more parsimonious models. They are particularly recommended when the data entail a large number of observations. That value of S is chosen for which the fit, as judged by these various criteria, is acceptable (i.e. select S which renders the minimum value). The major problem with the use of these criteria, and AIC in particular, is that they rely on the same asymptotic properties as the likelihood ratio test (Sclove, 1987), and can therefore only be used as indicative of the number of classes actually present. Recent work underway by Windham and Cutler (1991) using the estimated information matrix and the rates of convergence to estimate the number of classes appears as a promising alternative in this regard.

While the above heuristics account for overparameterization as large number of classes are derived, one must also ensure that the group centroids of the conditional densities are sufficiently separated for the solution that is selected. To assess the separation of the latent classes (for $S > 1$), an entropy measure can be utilized (see DeSarbo, Wedel, Vriens, and Ramaswamy, 1992) to examine the degree of fuzziness in latent class membership based on the estimated posterior probabilities:

$$E_S = 1 - \left[\sum_{i=1}^{n} \sum_{s=1}^{S} - \hat{\alpha}_{is} \ln \hat{\alpha}_{is}\right] / n \ln S \qquad (10.11)$$

E_S is a relative measure that is bounded between 0 and 1. A value of E_S close to zero, indicating that all the posteriors are equal for each observation, is of concern as it implies that the centroids of the conditional parametric distributions are not sufficiently well separated.

Latent Class Regression Models

There are a large number of applications in the physical and social sciences where the purpose of the analysis is the estimation of a linear model relating a dependent variable to a set of explanatory variables (see McCullagh and Nelder, 1989). In many of those applications however, the estimation of a single set of regression coefficients across all observations may be inadequate and potentially misleading if the observations arise from a number of (unknown) heterogeneous groups in which the coefficients differ. It is in these situations that latent class regression models have recently proven to be of great value. These latent class regression models, to be discussed in the sequel, are alternatively referred to as clusterwise regression models, a term originally coined by Späth (1979) for procedures that simultaneously cluster observations into a number of classes (using an exchange algorithm) and estimate regression models within each class (see also DeSarbo, Oliver, and Rangaswamy, 1990; Wedel and Kistmaker, 1989.

The literature on latent class regression models will be classified by the type of data to which the models are calibrated: normal distributed data, binary data, count data, constant sum data, or data arising from any member of the exponential family. Further, recently developed concomitant variable latent class regression models will be discussed. A taxonomy of the development of latent class regression models is provided in Table 10.1, where the order of presentation corresponds to the order in which the models are discussed by data-type below.

Latent class regression models for normal data

Work on regression mixtures was initiated by Quandt (1972), who introduced switching regression models, later extended by Hosmer (1974), and Quandt and Ramsey (1978). In switching regressions models, a linear function relating a univariate dependent variable $\mathbf{y}_i = (y_i)$ to P explanatory variables $\mathbf{x}_i = (x_{ip})$ $(i = 1, \ldots, n; p = 1, \ldots, P)$ is postulated:

$$y_i = \mathbf{x}_i \boldsymbol{\beta}_s + \varepsilon_i, \qquad (10.12)$$

Table 10.1 Applications of mixture regression models[a], [b]

Reference	*Mixture type*	*Estimation method*	Application
Normal data			
Quandt and Ramsey (1978)	Univariate normal	MD	Wage prediction
Quandt (1972)	Univariate normal	ML, NR	Housing construction
Goldfeld and Quandt (1973)	Univariate normal, hidden Markov	ML, NR	Housing construction
Cosslett and Lee (1985)	Univariate normal, hidden Markov regression	ML, NR	Cartel stability
Hamilton (1989)	Multivariate normal, hidden Markov time series	ML, NR	GNP-growth
Hamilton (1990)	Multivariate normal, hidden Markov time series	ML, EM	—
Hamilton (1991)	Multivariate normal, hidden Markov time series	QB, EM	Exchange rates
Engel and Hamilton (1990)	Multivariate normal, hidden Markov time series	ML, EM	Exchange rates
DeSarbo and Cron (1988)	Univariate normal	ML, EM	Trade show performance
Ramaswamy, DeSarbo, Reibstein and Robinson (1993)	Multivariate normal	ML, EM	Latent pooling for marketing Mix effects
DeSarbo, Wedel, Vriens and Ramaswamy (1992)	Multivariate normal	ML, EM	Metric conjoint analysis
Helsen, Jedidi, and DeSarbo (1992)	Multivariate normal	ML, EM	Country segmentation
Binary data			
Kamakura and Russell (1989)	Multinomial logit	ML, NR	Price segmentation
Kamakura (1992)	Multivariate multinomial logit	ML, NR	Value systems segmentation

Table 10.1 (*Contd.*)

Reference	*Mixture type*	*Estimation method*	Application
Bucklin and Gupta (1992)	Nested mult nomial logit	ML, NR	Purchase incidence and brand choice
Lwin and Martin (1989)	Binomial probit	ML, EM, NR	Parasite treatment resistance
De Soete and DeSarbo (1991)	Binomial probit	ML, EM	Choice of communication devices
Follmann and Lambert (1989)	Binomial logit, varying intercepts	ML, EM	Protozoan death rates
Wedel and DeSarbo (1992a)	Binomial logit	ML, EM	Paired comparison risk perception
Wedel and Leeflang (1992)	Binomial logit spline	ML, EM	Gabor Granger price experiments
Kamakura (1991)	Multinomial probit	ML, NR	External unfolding
Count data			
Wedel, DeSarbo, Bult, and Ramaswamy (1991)	Poisson	ML, EM	Direct mail address selection
Bucklin, Gupta, and Siddarth (1991)	Truncated poisson	ML, NR	Purchase frequency
Ramaswamy, Anderson, and DeSarbo (1993)	Negative Binomial	ML, EM	Purchase frequency
Constant sum data			
DeSarbo, Ramaswamy, and Chatterjee (1992)	Multivariate dirichlet	ML, EM	Multivariate constant sum conjoint analysis
Generalized linear model			
Wedel and DeSarbo (1992b)	Exponential family	ML, EM	Poisson regression coupon usage
Concomitant variable models			
Formann (1992)	Multinomial logit	ML, EM	Social mobility tables
Gupta and Chintagunta (1992)	Multinomial logit, probit and poisson	ML, NR	Scanner brand choice data

[a] ML = maximum likelihood; EM = expectation maximization algorithm; NR = Newton-Raphson algorithm; MD = minimum distance; QB = quasi-Bayes.

[b] Order of references conforms to their appearance in the text.

where the first component of $\mathbf{x}_i$ is a dummy variable taking the value 1, and the disturbance term ε_i is normally distributed with mean zero and variance σ^2. The parameters $\boldsymbol{\beta}_s = ((\beta_{ps}))$ take (for two "regimes") one of two unknown values for $s = 1$ and $s = 2$, depending on the unobserved "regime" that applies. The two unobserved regimes are represented by an underlying categorical variable denoting from which regime an observation arises. A two-component latent class regression model results. Goldfeld and Quandt (1973, 1976) and Cosslett and Lee (1985) develop hidden Markov switching regression models in which membership of observations in a regime is modelled by a Markov process. Hamilton (1989, 1990, 1991) and Engel and Hamilton (1990) extend the switching regression approach to time series models. The models describe discrete shifts in autoregressive parameters, where the shifts themselves are modeled by a hidden discrete-time Markov process. Whereas initially estimation was limited to small systems due to computational complexity involved in maximizing the likelihood (Hamilton, 1989), Hamilton (1990) proposed an EM algorithm that alleviates these problems by using the calculation of smoothed posterior probabilities in the E-step. In a third paper, Hamilton (1991) demonstrates advantages of the Quasi-Bayesian over the maximum likelihood approach for estimating the parameters, alleviating problems with potential singularities of the likelihood. The model was applied to the analysis of exchange rates (Engel and Hamilton, 1990; Hamilton, 1991). Titterington, Smith, and Makov (1985) review applications of switching regressions in economics.

DeSarbo and Cron (1988) first extended the stochastic switching regression models to more than two regimes. Assuming S component densities, with prior probabilities π_s $(s = 1, \ldots, S)$ as above, they formulate the conditional distribution of the dependent variable, given s, as:

$$f_{i|s}(y_i; \boldsymbol{\beta}_s, \sigma_s^2) = (2\pi\sigma_s^2)^{-1/2} \exp\left\{\frac{-(y_i - \mathbf{x}_i\boldsymbol{\beta}_s)^2}{2\sigma_s^2}\right\}, \tag{10.13}$$

which is conceptually similar to normal mixture approaches originally proposed by Wolfe (1970) and Day (1969), with the component means replaced by a linear predictor involving $\mathbf{x}_i$. The unconditional distribution follows from equation (10.3), and the likelihood from (10.6). (Unless the condition $\sigma_s^2 > 0$ is imposed, the likelihood is unbounded.) To estimate the parameters $\boldsymbol{\varphi}$ (where $\boldsymbol{\varphi}$ denotes the vector of all unknown parameters), the likelihood is maximized using the EM-algorithm, the development of which is analogous to that provided for unconditional mixtures in the previous section. Estimates of the posterior probabilities in the E-step, given provisional estimates $\boldsymbol{\varphi}^{(0)}$, are given by (10.4), and estimates of the prior probabilities are given by (10.8). In the M-step, closed form expressions for the estimates of the parameters $\boldsymbol{\beta}_s$ and σ_s^2 are:

$$\hat{\boldsymbol{\beta}}_s = \sum_{i=1}^{n} \hat{\alpha}_{is} (\mathbf{x}_i' x_i)^{-1} \sum_{i=1}^{n} \hat{\alpha}_{is} (\mathbf{x}_i' y_i), \tag{10.14}$$

$$\hat{\sigma}_s^2 = \sum_{i=1}^{n} \hat{\alpha}_{is} (y_i - x_i\hat{\beta}_s)^2 / \sum_{i=1}^{n} \hat{\alpha}_{is}. \tag{10.15}$$

DeSarbo and Cron provided a modest Monte Carlo analysis supporting the performance of the EM-algorithm under a variety of conditions. Whereas the above approach presents a univariate normal regression mixture, extensions to multivariate normal regression mixtures where repeated and correlated measures on each observational unit are present have also been developed (cf. Ramaswamy, DeSarbo, Reibstein, and Robinson, 1993; DeSarbo, Wedel, Vriens, and Ramaswamy, 1992; Helsen, Jedidi, and DeSarbo, 1992). Here, the conditional multivariate density of the dependent vectors $\mathbf{y}_i = ((y_{ij}))$, where j ($j = 1, \ldots, J$) indexes replications, given S, is:

$$f_{i|s}(y_i; \boldsymbol{\beta}_s, \boldsymbol{\Sigma}_s) = (2\pi)^{-1/2} |\boldsymbol{\Sigma}_s|^{-1/2} \exp\left\{-\frac{1}{2}(\mathbf{y}_i - \mathbf{x}_i\boldsymbol{\beta}_s)' \boldsymbol{\Sigma}_s^{-1} (y_i - x_i\boldsymbol{\beta}_s)\right\}, \tag{10.16}$$

where $\boldsymbol{\Sigma}_s$ denotes the variance–covariance matrix of $\mathbf{y}_i$ given class s. The unconditional density and likelihood are obtained from (10.3) and (10.6), estimates of the posterior and prior probabilities from (10.4) and (10.8). The EM algorithm is applied to maximize the likelihood for this model. Closed form expressions can be obtained for the parameters in the M-step as:

$$\hat{\boldsymbol{\beta}}_s = \left[\sum_{i=1}^{n} \hat{\alpha}_{is} (x_i' \boldsymbol{\Sigma}_s^{-1} x_i)\right]^{-1} \left[\sum_{i=1}^{n} \hat{\alpha}_{is} (\mathbf{x}_i' \boldsymbol{\Sigma}_s^{-1} \mathbf{y}_i)\right], \tag{10.17}$$

$$\hat{\boldsymbol{\Sigma}}_s = \sum_{i=1}^{n} \hat{\alpha}_{is} (\mathbf{y}_i - \mathbf{x}_i\hat{\boldsymbol{\beta}}_s)(\mathbf{y}_i - \mathbf{x}_i\hat{\boldsymbol{\beta}}_s)' / \sum_{i=1}^{n} \hat{\alpha}_{is}. \tag{10.18}$$

The above multivariate normal regression mixture has been applied to the analysis of conjoint measurement data (DeSarbo, Wedel, Vriens, and Ramaswamy, 1992), as a latent pooling method for simultaneously pooling and estimating linear regression models for cross-sectional-time-series data (Ramaswamy, DeSarbo, Reibstein, and Robinson, 1993), and to the analysis of diffusion patterns of consumer durable goods within segments of countries (Helsen, Jedidi, and DeSarbo, 1992). We later provide an application of this methodology to the measurement of service quality.

Latent class regression models for binary data

A mixture regression model for a multinomially distributed dependent variable was developed by Kamakura and Russell (1989). The dependent variable constitutes consumers' choices among a set of products or brands, where $y_{ijt} = 1$ if consumer i chooses brand j at time $t (j = 1, \ldots, J; t = 1, \ldots, T)$ and $y_{ijt} = 0$ otherwise. The independent variables are denoted by $\mathbf{x}_{ijt} = (x_{ijpt})$. The model is derived from random utility theory (McFadden, 1973), whereby conditional upon class s, the stochastic utility for a specific alternative j is formulated as a linear function of explanatory variables and an error term. The assumption that the error terms of the utility are distributed i.i.d. weibull leads to a multinomial logit model for the choices within class s, with conditional probabilities:

$$p_{ijt|s}(y_{ijt}; \boldsymbol{\beta}_s) = \frac{\exp(\mathbf{x}_{ijt}\boldsymbol{\beta}_s)}{\sum_{j=1}^{J} \exp(\mathbf{x}_{ijt}\boldsymbol{\beta}_s)}. \tag{10.19}$$

The conditional density is given by:

$$f_{i|s}(y_{ij}; \boldsymbol{\beta}_s) = \prod_{t=1}^{T} \prod_{j=1}^{J} p_{ijt|s}^{y_{ijt}}. \tag{10.20}$$

Expressions for the unconditional density, the likelihood, and posterior probabilities follow from (10.3), (10.5), (10.4). Kamakura and Russell maximized the likelihood for the multinomial mixture logistic regression model using a Newton-Raphson algorithm. The authors apply the latent class multinomial logit to the identification of segments of consumers that differ in price-sensitivity. Kamakura (1992) extended the latent class multinomial logit to a model for multiple dependent variables, assumed to be multinomially distributed and locally independent given segment s, which are modelled from a common set of explanatory variables. This model was applied to the analysis of value systems.

Another extension of the latent class multinomial logit was recently proposed by Bucklin and Gupta (1992) who proposed a latent class nested multinomial logit which models purchase incidence and brand choice simultaneously within S choice segments and R purchase incidence segments. The unconditional choice probabilities for this model are given by:

$$p_{ijt}(y_{ijt}, d_{it}; \boldsymbol{\varphi}) = \sum_{s=1}^{S} \sum_{r=1}^{R} \pi_{rs}\, p_{ijt|s}(y_{ijt}; \boldsymbol{\beta}_s)\, p_{it|r}(d_{it}; \mathbf{v}_r), \tag{10.21}$$

where the subscript t indicates time periods $(t = 1, \ldots, T)$, and $d_{it} = 1$ if a category purchase was made by household i at time t, and $y_{ijt} = 1$ if brand j was chosen by household i at time t. The term $p_{ijt|s}(\bullet)$ denotes the conditional probability that brand j is chosen, given a purchase of the product category at time t and segment s. This probability is modelled as a logit function of a set of explanatory variables, analogous to equation

(10.19). Here $p_{it|r}(\bullet)$ denotes the purchase incidence probability of the product category at time t, given purchase incidence segment r. This probability is modelled as a logit function of a set of explanatory variables and parameters $\mathbf{v}_r$, and an inclusive category value calculated from the brand choice model (Ben-Akiva and Lerman, 1985). The parameters of the model are estimated using a sequential procedure based on maximum likelihood. Initially households are classified into S segments on the basis of the brand choice model, and then the households are classified into R purchase incidence segments within each of the S choice segments, on the basis of the purchase incidence model. The model was applied to scanner panel data on household purchases of liquid detergents.

Lwin and Martin (1989), and De Soete and DeSarbo (1991) developed binomial mixture probit regression models for pick-any out of J data in which the conditional choice probabilities of $y_{ij} = 1$, indicating that subject i picks object j, given s, are given by:

$$p_{ijt|s}\,(y_{ijt};\, \boldsymbol{\beta}_s) = \Phi\,(\mathbf{x}_{ij}\, \boldsymbol{\beta}_s), \tag{10.22}$$

where $\Phi(\bullet)$ represents the cumulative normal distribution function. Follman and Lambert (1989) and Wedel and DeSarbo (1992a) propose mixtures of binomial logit regressions in which the choice probabilities are given by:

$$p_{ij|s}\,(y_{ijt};\, \boldsymbol{\beta}_s) = \frac{\exp(\mathbf{x}_{ij}\, \boldsymbol{\beta}_s)}{1 + \exp(x_{ij}\, \boldsymbol{\beta}_s)}. \tag{10.23}$$

(In the Follman and Lambert model, only the intercepts are allowed to vary across the S classes.) The conditional distribution of y_{ij} for the mixture logit and probit regression models is:

$$f_{i|s}\,(y_{ij};\, \boldsymbol{\beta}_s) = \prod_{j=1}^{J} p_{ij|s}\,(y_{ij};\, \boldsymbol{\beta}_s)^{y_{ij}}(1 - p_{ij|s}\,(y_{ij};\, \boldsymbol{\beta}_s))^{1 - y_{ij}}. \tag{10.24}$$

The unconditional distribution is given by (10.3), and the likelihood by (10.6). The above authors all use the EM-algorithm to maximize the likelihood. The estimates of the posterior probabilities in the E-step are given in (10.4). The estimates of $\boldsymbol{\beta}_s$ in the M-step are obtained from the stationary equations. These equations follow from (10.9) where the derivatives are now taken with respect to $\boldsymbol{\beta}_s$. Fisher's scoring method (De Soete and DeSarbo, 1991) or the method of iterative reweighted least-squares (Wedel and DeSarbo, 1992a) are used to solve these equations. Lwin and Martin (1989) employed a Newton-Raphson procedure to maximize the likelihood across the entire parameter space. Wedel and DeSarbo (1992a) perform a small Monte Carlo study which supports the performance of the algorithm under a variety of data conditions. The binomial mixture regression models have been applied to the analysis of the effects of a poison on the death-rate of a protozoan trypanosome (Follmann and Lambert, 1989), to the analysis of the resistance to treatment of parasites of sheep as a function of the treatment dose (Lwin and Martin, 1989), to data on consumers' choices of communication devices as a function of the attributes of such devices (De Soete and DeSarbo, 1991), and to the analysis of paired comparison choices of consumers' indicating perceived risk with respect to automobiles, as a function of the attributes of these automobiles (Wedel and DeSarbo, 1992a). Wedel and Leeflang (1992) extended the mixture binomial logit of Wedel and DeSarbo (1992a) to include regression splines, and applied their model to the analyses of consumers (binary) purchase responses to incremental price-levels in so called Gabor Granger price experiments.

Kamakura (1991) proposed a latent class multinomial probit regression model. Formulation and estimation of this model, using a Newton-Raphson procedure, is analogous to that of the Kamakura and Russell (1989) latent class multinomial logit model, with expression (10.19) replaced by a multivariate probit function. The latent class multinomial probit regression procedure was used to estimate a mixture of (external) ideal-point models on data on pairwise choices among a number of stimuli.

Latent class regression models for count data

Wedel, DeSarbo, Bult, and Ramaswamy (1991) proposed a latent class poisson regression, Bucklin, Gupta, and Siddarth (1991) a latent class truncated poisson regression, and Ramaswamy, Anderson, and DeSarbo (1993) a latent class negative binomial regression model. These models assume that the dependent variable y_{ijt} constitutes an integer count of, say, a number of products purchased by consumer i at time period t, following a poisson, a truncated poisson, or a negative binomial process respectively. Conditional upon class s, the expectation of these observations is modelled as:

$$\lambda_{it|s} = \exp\{\mathbf{x}_{it}\,\boldsymbol{\beta}_s\}. \tag{10.25}$$

The conditional densities for the poisson, truncated poisson, and negative binomial are then respectively:

$$f_{i|s}(y_{it}, t_{it}; \boldsymbol{\beta}_s) = \prod_{t=1}^{T} \frac{\exp\{-\lambda_{it|s}\, t_{it}\}(\lambda_{it|s}\, t_{it})^{y_{it}}}{y_{it}\,!}, \tag{10.26}$$

$$f_{i|s}(y_{it}; \boldsymbol{\beta}_s) = \prod_{t=1}^{T} \frac{\exp\{-\lambda_{it|s}\}(\lambda_{it|s})^{y_{it}}}{(1-\exp\{-\lambda_{it|s}\}\, y_{it}\,!}, \tag{10.27}$$

$$f_{i|s}(y_{it}; \boldsymbol{\beta}_s) = \prod_{t=1}^{T} \frac{\Gamma(\delta + y_{it})}{y_{it}\,!\,\Gamma(\delta)} \left[\frac{\delta}{\delta + \lambda_{it|s}}\right] \left[\frac{\lambda_{it|s}}{\delta + \lambda_{it|s}}\right]^{y_{it}}, \tag{10.28}$$

where t_{it} denotes the length of time period t for subject i which is included as an offset in the poisson model, δ denotes the precision parameter of the NBD, and $\Gamma(\bullet)$ the gamma function. The truncated poisson accommodates non-zero positive counts, and the negative binomial accommodates heterogeneity within classes through the precision parameter. The likelihood of these models is obtained from (10.3) and (10.6), the posterior probabilities from (10.4). Whereas Bucklin, Gupta, and Siddarth (1991) use a Newton-Raphson method to maximize the likelihood over the entire parameter space, Wedel, DeSarbo, Bult, and Ramaswamy (1991), and Ramaswamy, Anderson, and DeSarbo (1993) use an EM-algorithm, the development of which follows from equations (10.7), (10.8) and (10.9), where the derivatives in (10.9) are taken with respect to $\boldsymbol{\beta}_s$ (respectively δ). Bucklin, Gupta, and Siddarth (1991) and Ramaswamy, Anderson, and DeSarbo (1993) apply their models to scanner data on purchase frequencies of non-durable goods, while Wedel, DeSarbo, Bult, and Ramaswamy (1991) apply their model to the selection of customers from direct mail databases.

A latent class regression model for constant sum data

For the analysis of multivariate constant-sum data, DeSarbo, Ramaswamy, and Chatterjee (1992) proposed a latent class dirichlet regression model. In this model, K dependent variables, with J replications per observational unit each, are assumed to be dirichlet, with parameters θ_{jks}, conditional upon segment s. These θ_{jks} are reparameterized in terms of a common set of explanatory variables $\mathbf{x}_j = (x_{jp})$, and a set of coefficients $\boldsymbol{\beta}_{ks} = (\beta_{kps})$:

$$\theta_{jks} = \exp\{\mathbf{x}_j \boldsymbol{\beta}_{ks}\}. \tag{10.29}$$

The conditional density, given s, is:

$$f_{i|s}(y_{ijk}; \boldsymbol{\beta}_{ks}) = \prod_{j=1}^{J} \frac{\Gamma\left(\sum_{k=1}^{K} \theta_{jks}\right)}{\prod_{k=1}^{K} \Gamma(\theta_{jks})} \prod_{k=1}^{K} y_{ijk}^{\alpha_{jks}-1}. \tag{10.30}$$

The unconditional distribution and the likelihood follow from (10.3) and (10.6). The authors estimate the model using the EM-algorithm, where estimates of the posteriors in the E-step are obtained from (10.4), and in the M-step the estimates of the priors are obtained from (10.8), the estimates of $\boldsymbol{\beta}_{ks}$ are obtained from the stationary equations (10.9), with the derivatives taken with respect to $\boldsymbol{\beta}_{ks}$. DeSarbo, Ramaswamy, and Chatterjee (1992) use a gradient based search to solve these stationary equations in the M-step. The model was applied to a conjoint study on industrial purchasing in which a sample of managers provided constant sum ratings of profiles of product-types on a set of supplier selection criteria.

A general latent class regression model for data from the exponential family

Wedel and DeSarbo (1992b) propose a generalized linear regression mixture model. This model handles many of the above existing mixture regression procedures as special cases, as well as a host of other parametric specifications in the exponential family heretofore not mentioned in the latent class regression literature. Conditional upon class s, a generalized linear model (McCullagh and Nelder, 1989) is formulated consisting of a specification of the distribution of the dependent variable, y_i, as one of the exponential family, a linear predictor, η_{is}, and a function $g(\bullet)$, which links the random and systematic components. The conditional probability density function of y_i, given class s, takes the form:

$$f_{i|s}(y_i \mid \theta_{is}, \lambda_s) = \exp\{y_i \theta_{is} - b(\theta_{is}))/a(\lambda_s) + c(y_i, \lambda_s)\}, \qquad (10.31)$$

for specific functions $a(\bullet)$, $b(\bullet)$ and $c(\bullet)$, where conditional upon class s, the y_i are i.i.d. with canonical parameters θ_{is} and means μ_{is}. The parameter λ_s is a dispersion parameter, and is assumed to be constant over observations in class i, while $a(\lambda_s) > 0$. The link function $g(\bullet)$ is defined as:

$$\eta_{is} = g(\mu_{is}), \qquad (10.32)$$

where the linear predictor is produced by the covariates $\mathbf{x}_i$ and the parameter vectors $\boldsymbol{\beta}_s$ in class s:

$$\eta_{is} = \mathbf{x}_i \boldsymbol{\beta}_s. \qquad (10.34)$$

So called canonical links occur when $\theta_{ij} = \eta_{ij}$ (respectively the identity, log, logit, inverse, and squared inverse functions for the normal, poisson, binomial, gamma, and inverse gaussian distributions; see McCullagh and Nelder, 1989). An estimate of the parameters is obtained by maximizing the likelihood equation, obtained from (10.6). The authors show that this can be done using an EM-algorithm. In the E-step, estimates of the posterior probabilities are obtained from (10.4). In the M-step, estimates of the prior probabilities are obtained from (10.8), while estimates of the regression parameters are obtained from the stationary equations (10.9), where the derivatives are taken with respect to the parameters $\boldsymbol{\beta}_s$ and λ_s. The stationary equations are:

$$\sum_{i=1}^{n} \hat{\alpha}_{is}^{(0)} V_{is} (y_i - \mu_{is}) x_{ip} \frac{d\mu_{is}}{d\eta_{is}} = 0. \qquad (10.34)$$

Equation (10.34) is the ordinary stationary equation of a generalized linear model fitted across all observations, where observation j contributes to the estimating equations with fixed weights $\hat{\alpha}_{is}^{(0)}$. Wedel and DeSarbo propose to solve the stationary equations in the M-step, for each class s, using the iterative reweighted least squares procedure proposed by Nelder and Wedderburn (1972) for ML estimation of generalized linear models, with each observation weighted with $\hat{\alpha}_{is}^{(0)}$. The procedure was illustrated by Wedel and DeSarbo (1992b) with an application to the analysis of consumers' coupon usage.

An Application of Latent Class Regression to Conjoint Analysis: Service Quality Measurement

The early exploratory research of Parasuraman, Zeithaml, and Berry (1985) revealed that the primary criteria utilized by consumers in assessing service quality can be described by some ten separate dimensions:

1 tangibles
2 reliability
3 responsiveness
4 communication
5 credibility
6 security
7 competence
8 courtesy
9 understanding/knowing the consumer
10 access.

Table 10.2 Parasuraman, Zeithaml, and Berry (1985) determinants of service quality

Reliability involves consistency of performance and dependability.
It means that the firm performs the service right the first time.
It also means that the firm honors its promises. Specifically, it involves:
- accuracy in billing;
- keeping records correctly;
- performing the service at the designated time.

Responsiveness concerns the willingness or readiness of employees to provide service. It involves timeliness of service:
- mailing a transaction slip immediately;
- calling the customer back quickly;
- giving prompt service (e.g. setting up appointments quickly).

Competence means possession of the required skills and knowledge to perform the service. It involves:
- knowledge and skill of the contact personnel;

Table 10.2 *(Contd.)*

- knowledge and skill of operational support personnel;
- research capability of the organization, e.g. securities brokerage firm.

Access involves approachability and ease of contact. It means:
- the service is easily accessible by telephone (lines are not busy and they don't put you on hold);
- waiting time to receive service (e.g. at a bank) is not extensive;
- convenient hours of operation;
- convenient location of service facility.

Courtesy involves politeness, respect, consideration
- consideration for the consumer's property (e.g. no muddy shoes on the carpet);
- clean and neat appearance of public contact personnel.

Communication means keeping customers informed in language they can understand and listening to them. It may mean that the company has to adjust its language for different consumers – increasing the level of sophistication with a well-educated customer and speaking simply and plainly with a novice. It involves:
- explaining the service itself;
- explaining how much the service will cost;
- explaining the trade-offs between service and cost;
- assuring the consumer that a problem will be handled.

Credibility involves trustworthiness, believability
- company name;
- company reputation;
- personal characteristics of the contact personnel;
- the degree of hard sell involved in interactions with the customer.

Security is the freedom from danger
- physical safety (Will I get mugged at the automatic teller machine?);
- financial security (Does the company know where my stock certificate is?);
- confidentiality (Are my dealings with the company private?).

Understanding/knowing the customer involves making the effort to understand the customer's needs. It involves:
- learning the customer's specific requirements;
- providing individualized attention;
- recognizing the regular customer.

Tangibles include the physical evidence of the service: physical facilities;
- appearance of personnel;
- tools or equipment used to provide the service;
- physical representations of the service
- other customers in the service facility.

Source Taken from Parasuraman, Zeithaml, and Berry (1985).

Table 10.2 presents the actual description of these ten dimensions from these authors which serves as the essential structure of the service-quality domain from which specific items were derived for the SERVQUAL instrument.

In its present form (see Parasuraman, Zeithaml, and Berry, 1988), SERVQUAL contains 22 pairs of Likert-type items. One set of measures, containing one item from each pair, is utilized to measure customers' expected levels of services for a particular service industry as a way of calibrating expectations. The second remaining set of measures, containing the other item from each pair, is intended to measure customers' perceived level of service provided by a specific service company as a way of calibrating perceptions. An aggregate measure of service quality is then formulated by summing the difference scores between the corresponding set of items (i.e. perceptions minus expectations).

Since this important work, several authors have criticized the use of this instrument in actual applied settings. Carman (1990) suggested that the number and type of dimensions may vary by service category. He also found problems in attempting to use the same wording across different service categories, as well as ambiguity in dealing with services that provide multiple service functions (e.g hospitals). Another valid criticism of Carman (1990) concerns the analysis of the difference scores between perceptions and expectations collected separately in the SERVQUAL framework and his questioning of the psychometric properties of such a difference scale. Babakus and Boller (1992) have also criticized the SERVQUAL approach in that the number of dimensions is likely to depend upon the service category under study. These authors also question the use of difference scores formed from subtracting expectations from perceptions. The mixed wording utilized in SERVQUAL, according to Babakus and Boller (1992), may also lead to superficial method factors in subsequent analyses. Cronin and Taylor (1992) more recently claim that the SERVQUAL instrument confounds the measurement of service satisfaction with service quality. In the empirical work these authors performed, the individual SERVQUAL item reliabilities and the convergent and discriminant validity of the measures were found questionable.

DeSarbo, Huff, Rolandelli, and Choi (1993) recently devised an alternative measurement scheme for the measurement of perceived service quality based on conjoint analysis (Green and Rao, 1971) that can be easily modified to any service category. They propose the use of this measurement procedure in the general SERVQUAL framework utilizing service-specific category operationalizations of the 10 dimensions originally proposed by Parasuraman, Zeithaml, and Berry (1988) in an expectancy confirmation/disconfirmation response manner. The advantages of this procedure are (1) it measures true perceptions, as opposed to perceptions confounded with expectations and satisfaction as in the Parasuraman, Zeithaml and Berry approach; (2) the number, type, and operationalization of the specific dimensions (*vis- à-vis* the wording) are completely flexible according to the specific usage scenario; (3) estimation can be performed in an efficient manner utilizing orthogonal designs and simple OLS; (4) the proposed model can lead to interesting quality optimization models, as well as models that explore segmentation. DeSarbo *et al.* (1993) illustrate this approach with respect to service quality perceptions of banks and dental offices. We will briefly review their research design for service quality evaluation of banks, and apply the DeSarbo, Wedel, Vriens, and Ramaswamy (1992) latent class conjoint regression methodology to this data to explore sample heterogeneity in terms of market segments.

Study design

Based on several in-depth interviews with students and an extensive literature review (e.g. Parasuraman, Zeithaml, and Berry (1988) have previously examined banks), DeSarbo *et al.* (1993) generated tailor-made operationalizations of the original set of 10 SERVQUAL *dimensions for banks shown in Table 10.3. The authors chose to* split the tangibles factor into two separate variables for banks due to the (*a priori* determined) importance and complexity of this factor, resulting in 11 conjoint factors. In the absence of prior theory, and to reduce respondent fatigue, a 3^{11} fractional factorial design was selected for main effects-only estimation (see Addelman, 1962).

Table 10.3 SERVQUAL factors utilized in conjoint experiment

	Banks		SERVQUAL *dimensions*
A	Facility and equipment	}	
B	Selection and quality of financial offerings	}	⇐ *Tangibles*
C	Accuracy and dependability		⇐ *Reliability*
D	Speed of service		⇐ *Responsiveness*
E	Communication with customers		⇐ *Communication*
F	Reputation for honesty and integrity		⇐ *Credibility*
G	Financial strength and security		⇐ *Security*
H	Knowledge and competence of personnel		⇐ *Competence*
I	Politeness and courtesy of personnel		⇐ *Courtesy*
J	Understanding of individual customer needs		⇐ *Understanding*
K	Convenience of location and operating hours		⇐ *Access*

The actual design matrix, converted to dummy variables, is shown in Table 10.4. Twenty-seven profiles were used for estimation, and the last three profiles for validation. Note that the levels of each factor are coded so that "same as expected" is always represented by (0,0), while "worse than expected" is represented by (1,0), and "better than expected" is represented by (0,1). This was purposely done in order to directly estimate possible asymmetric effects between the positive and negative level states for each of the factors. The estimates for the intermediate or neutral level, "same as expected," for each factor are thus confounded with the intercept term.

After two rounds of pretesting, a questionnaire was developed for banks. Respondents were first asked to list the attributes s/he thought were important in their use of banks in an open ended framework. They were then asked to evaluate the importance of the 11 factors described in Table 10.3 on a 9-point scale. The conjoint task then followed where the 30 profiles were randomized, as well as the order of the factors. The authors then asked a battery of questions concerning the respondents' evaluations of their current bank including an overall quality assessment, ratings of their bank's performance on these 11 factors based on their previous expectations, and usage and experience levels with the particular aspects of banks. Finally, demographic questions concerning age, marital status, gender, home ownership, and level of education were included in the survey. Fifty-three students completed this bank questionnaire.

The conjoint model and aggregate OLS results

DeSarbo *et al.* (1993) focus on a standard main-effects part-worth model estimated by ordinary least squares (OLS). The response of a given respondent to the j^{th} profile is given by:

$$y_j = \sum_{p=1}^{P} \sum_{q=1}^{Q_p} \beta_{pq} x_{jpq} + \varepsilon_j, \tag{10.35}$$

where:

y_j = the perceived service quality judgment for the j^{th} experimental profile ($j = 1, \ldots, 30$);
β_{pq} = the part-worth of the q^{th} level of the p^{th} SERVQUAL factor;
x_{jpq} = a dummy variable that has the value of 1 if profile j takes on the q^{th} level of the p^{th} SERVQUAL factor, and zero otherwise;
Q_p = the number of levels of the p^{th} SERVQUAL factor (here $Q_p = 3$ for all p)
p = the number of SERVQUAL factors ($p = 11$), and
ε_j = an error term.

With $J(= 30)$ profiles, the relationships in (10.35) for a given respondent can be summarized in matrix form via:

$$\mathbf{Y} = \mathbf{XB} + \mathbf{U}, \tag{10.36}$$

where:

$\mathbf{Y}' = (y_1, y_2, \ldots, y_J)$;
$\mathbf{X}$ = a $J \times K$ dummy variable matrix with a column of 1's and $q_p - 1$ columns to code a factor with q_p levels;

$$K = \sum_{p=1}^{P} Q_p - p + 1;$$

Table 10.4 3^{11} fractional factorial design with validation profiles

Profile	A		B		C		D		E		F		G		H		I		J		K	
	SERVQUAL *dimension/factor*																					
1	1	0	1	0	1	0	1	0	1	0	1	0	1	0	1	0	1	0	1	0	1	0
2	1	0	1	0	1	0	1	0	0	0	0	0	0	1	0	0	0	1	0	0	0	1
3	1	0	1	0	1	0	1	0	0	1	0	1	0	0	0	1	0	0	0	1	0	0
4	1	0	0	0	0	0	0	1	1	0	1	0	1	0	0	0	0	0	0	0	0	0
5	1	0	0	0	0	0	0	1	0	0	0	0	0	1	0	1	1	0	0	1	1	0
6	1	0	0	0	0	0	0	1	0	1	0	1	0	0	1	0	0	1	1	0	0	1
7	1	0	0	1	0	1	0	0	1	0	1	0	1	0	0	1	0	1	0	1	0	1
8	1	0	0	1	0	1	0	0	0	0	0	0	0	1	1	0	0	0	1	0	0	0
9	1	0	0	1	0	1	0	0	0	1	0	1	0	0	0	0	1	0	0	0	1	0
10	0	0	1	0	0	0	0	0	1	0	0	0	0	0	1	0	1	0	0	0	0	0
11	0	0	1	0	0	0	0	0	0	0	0	1	1	0	0	0	0	1	0	1	1	0
12	0	0	1	0	0	0	0	0	0	1	1	0	0	1	0	1	0	0	1	0	0	1
13	0	0	0	0	0	1	1	0	1	0	0	0	0	0	0	0	0	0	0	1	0	1
14	0	0	0	0	0	1	1	0	0	0	0	1	1	0	0	1	1	0	1	0	0	0
15	0	0	0	0	0	1	1	0	0	1	1	0	0	1	1	0	0	1	0	0	1	0
16	0	0	0	1	1	0	0	1	1	0	0	0	0	0	0	1	0	1	1	0	1	0
17	0	0	0	1	1	0	0	1	0	0	0	1	1	0	1	0	0	0	0	0	0	1
18	0	0	0	1	1	0	0	1	0	1	1	0	0	1	0	0	1	0	0	1	0	0
19	0	1	1	0	0	1	0	1	1	0	0	1	0	1	1	0	1	0	0	1	0	1
20	0	1	1	0	0	1	0	1	0	0	1	0	0	0	0	0	0	1	1	0	0	0

Table 10.4 (*Contd*).

	SERVQUAL *dimension/factor*																					
Profile	*A*		*B*		*C*		*D*		*E*		*F*		*G*		*H*		*I*		*J*		*K*	
21	0	1	1	0	0	1	0	1	0	1	0	0	1	0	0	1	0	0	0	0	1	0
22	0	1	0	0	1	0	0	0	1	0	0	1	0	1	0	0	0	0	1	0	1	0
23	0	1	0	0	1	0	0	0	0	0	1	0	0	0	0	1	1	0	0	0	0	1
24	0	1	0	0	1	0	0	0	0	1	0	0	1	0	1	0	0	1	0	1	0	0
25	0	1	0	1	0	0	1	0	1	0	0	1	0	1	0	1	0	1	0	0	0	0
26	0	1	0	1	0	0	1	0	0	0	1	0	0	0	1	0	0	0	0	1	1	0
27	0	1	0	1	0	0	1	0	0	1	0	0	1	0	0	0	1	0	1	0	0	1
28	0	0	1	0	0	1	0	0	0	0	1	0	0	1	1	0	0	0	1	0	1	0
29	0	1	1	0	0	0	0	0	0	0	0	0	0	0	1	0	0	1	0	1	0	1
30	1	0	0	0	0	1	0	1	0	1	1	0	1	0	0	1	0	1	0	1	0	1

Table 10.5 Aggregate conjoint results (OLS) for banks

		Aggregate Coefficient	*Factor Importance*
Intercept		4.46***	
A. Equipment:	Worse	−0.13*	0.32
	Better	0.19**	
B. Offering/ Office:	Worse	−0.44***	0.60
	Better	0.16**	
C. Dependability:	Worse	−0.60***	0.85
	Better	0.25***	
D. Speed:	Worse	−0.28***	0.31
	Better	0.03	
E. Communication:	Worse	−0.01	0.16
	Better	0.15**	
F. Integrity:	Worse	−0.60***	0.80
	Better	0.20***	
G. Security:	Worse	−0.58***	0.66
	Better	0.08	
H. Competence:	Worse	−0.28***	0.28
	Better	0.00	
I. Courtesy:	Worse	−0.27***	0.36
	Better	0.09	
J. Understanding:	Worse	−0.09	0.13
	Better	0.04	
K. Access:	Worse	−0.30***	0.48
	Better	0.18***	
R^2		0.29	
F		26.06***	

*$p \leq 0.10$
**$p \leq 0.05$
***$p \leq 0.01$

$\mathbf{B}' = (\beta_0, \beta_1, \ldots, \beta_{K-1})$;
$E' = \varepsilon_1, \varepsilon_2, \ldots, \varepsilon_J)$;

and the prime denotes transpose. The estimate of $\mathbf{B}'$ will be denoted $\mathbf{b}' = (b_0, b_1, \ldots, b_{K-1})$ and is equal to $(\mathbf{X}'\,\mathbf{X})^{-1}\,\mathbf{X}'\,\mathbf{Y}$, where the negative one signifies a matrix inverse. In our application, $\mathbf{X}$ corresponds to an orthogonal fractional factorial design (e.g. Addelman, 1962). Note that expressions (10.35) and (10.36) can be specified and estimated by respondent (individual level analysis) or over the entire sample via an aggregate or pooled analysis.

Table 10.5 presents the multiple regression results for the aggregate (over all subjects), pooled sample for banks. The table also portrays the importances for each of the SERVQUAL dimensions. (As in conjoint analysis, these importances were calculated on the basis of the range of the coefficients for the three levels within each SERVQUAL factor.) With very few exceptions, the absolute value of the coefficients in Table 10.5 for "worse than expected" levels are much greater than those for "better than expected" levels for both services across most SERVQUAL dimensions reflecting an interesting asymmetry in the responses. This implies that the costs induced by not meeting customers expectations (negative disconfirmation) may exceed the benefits of exceeding those expectations (positive disconfirmation) – a result also found by Oliver and DeSarbo (1988) and Anderson and Sullivan (1993) in a satisfaction context. Such results may also be explained through prospect theory (Kahneman and Tversky, 1979) via a risk-aversion tendency on the part of the majority of respondents. For respondents who may be characterized as seeking minimal variance (deviation from expectation) and maximal mean net benefit (expected amount of net benefit), experiencing the negative factor level (worse than expected) will both decrease the mean and increase the variance, while experiencing the positive factor level (better than expected) will increase the mean while increasing the variance. This argument implies, therefore, that the "worse than expected" levels would have a much larger impact on responses than the "better than expected" levels, as witnessed in Table 10.5. For banks, dependability and integrity are the most important SERVQUAL-based dimensions in assessing perceptions of service quality, followed by security and financial offerings.

Latent class conjoint analysis

Table 10.6 presents the various goodness-of-fit indices available in DeSarbo *et al.* (1992) latent class conjoint analysis procedure as applied to this particular bank service quality data set. As seen, the $S = 3$ latent class solution appears to have the minimum AIC value, with entropy of 0.958 suggesting sufficient centroid separation among the three multivariate normal conditional distributions.

Table 10.6 Latent class conjoint analysis results

S	*LnL*	*Number of parameters*	AIC	*Entropy*
1	−1920.126	50	3940.252	0.000
2	−1857.391	101	3916.783	0.883
3	−1805.640	152	3915.280[a]	0.958
4	−1769.074	203	3944.148	0.983
5	−1705.439	254	3918.878	0.991

[a] Denotes minimum AIC.

Table 10.7 presents the part-worth regression coefficients, factor importance, and mixing proportions by latent class for the $S = 3$ solution. As with the aggregate OLS solution presented in Table 10.5, we also see the same

Table 10.7 Latent class conjoint results for S = 3 market segments

			Latent Class 1		Latent Class 2		Latent Class 3	
			Coefficient	*Importance*	*Coefficient*	*Importance*	*Coefficient*	*Importance*
Intercept			0.08		0.87		0.13	
A	Equipment:	Worse	−0.36	0.53*	−0.54	0.63*	−0.38	0.74*
		Better	0.17		0.09		0.36	
B	Offering/	Worse	−0.11	0.31	−0.18	0.46	0.03	0.22
	Office:	Better	0.20		0.28		0.19	
C	Dependability:	Worse	−0.44	0.58*	−1.07	1.48*	−0.38	0.73*
		Better	0.14		0.41		0.35	
D	Speed:	Worse	−0.54	0.54*	0.03	0.03	−0.07	−0.21
		Better	−0.05		0.03		0.14	
E	Communication:	Worse	−0.07	0.21	0.13	0.03	−0.02	0.31
		Better	0.14		0.03		0.29	
F	Integrity:	Worse	−0.44	0.70*	−0.99	1.11*	−0.38	0.42
		Better	0.26		0.21		0.04	
G	Security:	Worse	−0.77	0.87*	−0.67	0.89*	0.03	0.08
		Better	0.10		0.22		−0.05	
H	Competence:	Worse	−0.17	0.21	−0.24	0.24	−0.52	0.55*
		Better	0.04		−0.09		0.03	
I	Courtesy:	Worse	−0.25	0.25	−0.20	0.20	−0.49	0.99*
		Better	−0.01		−0.04		0.50	
J	Understanding:	Worse	−0.12	0.21	−0.02	0.02	−0.13	0.13
		Better	0.09		0.00		−0.01	
K	Access:	Worse	−0.40	0.56*	0.05	0.05	−0.38	0.86*
		Better	0.16		0.05		0.48	
λ			0.52		0.27		0.21	

*indicates factor importances > 0.50.

pattern of minus and plus signs for the larger derived coefficients (worse, better) across the three latent classes. However, a cursory inspection of the factor importances reveals substantial heterogeneity in the sample. Latent class one (52% of the sample) members find security and integrity as the most significant drivers of service quality, perhaps reflecting a *conservative* concern for security. Latent class two (27% of the sample) members find dependability and integrity, as the most important factors affecting their perceptions of bank service quality, perhaps reflecting a concern for *consistent* performance. Finally, latent class three (21% of the sample) members find courtesy and access as the most important factors affecting their perceptions of service quality perhaps reflecting *convenience and personal interactions*. The table also reveals the other important factors by latent class.

As in the aggregate case, the magnitude of the coefficients reflecting the "worse than" condition as uniformly larger across the three latent classes for the most part than the coefficients for the "better than" condition indicating the asymmetry in the two conditions. However, a cursory comparison of Tables 10.5 and 10.7 illustrates how an aggregate or pooled analysis can mask interesting heterogeneity in a sample of consumers which could potentially mislead the market researcher. Note, further analysis is possible here by relating the derived posterior probabilities to individual background characteristics (e.g. demographics) in order to make the derived segements more accessible.

Recent Developments

Latent class regression models are important developments of mixture models for classification for use in marketing given the importance of market segmentation. Recently, Titterington (1990) identified some recent innovative work in this field. He notes that the traditional mixture model as outlined above, implies a "hidden" multinomial, in the sense that the random variables representing class membership are assumed to be independently distributed according to an S-cell multinomial distribution, as was previously indicated. An important extension of this notion pertains to models for sequential data in which the unobserved variable $z_{is}(i = 1, \ldots, n)$ are assumed to follow a stationary Markov chain with a finite, but unknown, number of latent states S. For the hidden Markov mixture the complete-data log-likelihood can be formed based on independence of $\mathbf{Y}$ and $\mathbf{Z}$, and the EM-algorithm can be used to estimate the parameters (see Titterington, Smith and Makov, 1985). The E-step, however, requires backward and forward recursions (Levinson, Rabiner, and Sondhi, 1983; Pickett and Whiting, 1987) to estimate the conditional expectations of the unobserved data. The M-step can generally be explicitly formulated. Developments on hidden Markov models have been published by Baum and Eagon (1967), Baum *et al.* (1970), and Lindgren (1978), as well as Hamilton (1989, 1990, 1991) and Engel and Hamilton (1990). Hidden Markov models for discrete manifest data formulated in a latent class context have been proposed by Wiggins (1973), Poulsen (1982, 1990), van de Pol and de Leeuw (1986), and Bye and Schechter (1986). These models have been applied in the context of speech recognition (Levinson, Rabiner, and Sondhi, 1983), the analysis of brand-switching behavior (Poulsen, 1990), and exchange rates (Engel and Hamilton, 1990).

Hidden Markov random field (MRF) models are spatial analogues of hidden Markov chains. Here the unobserved variables indicating class membership are dependent on neighborhoods defined by cliques, and are assumed to follow a gibbs distribution. As both the E-and M-steps of the EM-algorithm for maximizing the complete log-likelihood for hidden MRF models are numerically complex, a number of other methodologies for estimation have been developed (see Titterington, 1990, for a review). MRF models have been used as models for image analysis (Geman and Geman, 1984; Besag, 1986).

The recent developments in latent class models, especially in psychometric applications, are numerous and a complete review is beyond the scope of this chapter. Comprehensive reviews have been given by Clogg (1981), Andersen (1982), Langeheine (1988), and Dillon (1993). DeSarbo, Manrai, and Manrai (1993) have reviewed latent class multidimensional scaling models in this handbook. Considerable interest appears to have been devoted to latent class models in which constraints are imposed upon the parameters. The models developed subsume categorical data analogs of linear structural relations models (Goodman 1974a,b; Clogg, 1981; Formann, 1982, 1984, 1985), Scaling models (MacReady and Dayton, 1980; Rindskopf, 1983; Haertel, 1984), and mixed Markov models (Poulsen, 1982, 1990; van de Pol and Langeheine, 1989). Habermans' (1979) formulation appears to be the most general (Langeheine, 1988), and is equivalent to that of a log-linear model for frequencies in which unobserved variables are included. Mooijaart and van der Heijden (1992) indicate some problems related to the application of the EM-algorithm for estimating constrained latent class models.

Dayton and MacReady (1988) proposed a model that imposes a specific structure on the prior probabilities in latent class models for multivariate categorical data. They formulate a so-called submodel in which the conditional relation between external concomitant variables and prior probabilities is modelled. The model was named a concomitant variable latent class model. For L concomitant variables u_{il} ($i = 1, \ldots, n$; $1 = 1, \ldots, L$), the submodel takes the general form:

$$\pi_{is} = h\left(\sum_{l=1}^{L} \gamma_{ls}\, u_{il}\right), \qquad (10.37)$$

for some function $h(\bullet)$, which preserves the constraints on the π_{is}, such as the logistic function. Dayton and MacReady estimate the concomitant variable latent class model using the simplex method. A number of authors have recently extended the Dayton and MacReady (1988) approach to tent class regression models. Formann (1992), Gupta and Chintagunta 992), and Kamakura, Wedel, and Agrawal (1992) propose latent class ultinomial logit regression models, with conditional choice probabilities ırmulated as a function of explanatory variables, as in equation (10.19). illon, Kumar, and de Borrero (1993) recently develop a latent class extended TL model for capturing individual differences in paired comparisons ata. Prior probabilities are formulated as a function of a set of concomi- ınt variables, as in equation (10.11), where the function $h(\bullet)$ is taken to e the logit function. Gupta and Chintagunta (1992) also develop related ncomitant variable latent class regression model in which the dependent ariable is assumed to follow a truncated poisson, as well as a concomitant ariable latent class multinomial probit regression model. All these models multaneously identify latent classes, estimate regression models within ach class, and estimate the relationship of class membership with con- omitant variables. Gupta and Chintagunta (1992) and Kamakura, Wedel, nd Agrawal (1992) use a Newton-Raphson procedure to maximize the kelihood over the entire parameter space. Formann (1992) develops an M algorithm for the estimation of the parameters which involves the olution of two independent sets of nonlinear equations in the M-step – ne for the conditional probability regression parameters $\boldsymbol{\beta}_s$, and one for ıe prior probability regression parameters, γ_{ls}. Dillon *et al.* (1993) also mploy the EM-algorithm to estimate their model. Applications of these ıodels to brand choice, and quantity purchased are provided by Gupta nd Chintagunta, and Kamakura and Agrawal, while Formann presents everal applications among which are the analysis of social mobility tables, nd paired comparison data.

)uestions

In the application of the latent class conjoint segmentation procedures of DeSarbo, Wedel, Vriens, and Ramaswamy (1992) to the service quality data, describe how the expectation of a preference rating y_j can be obtained.

Discuss the major advantages of using the latent class regression procedures such as those proposed by DeSarbo, Wedel, Vriens, and Ramaswamy (1992), Kamakura and Russell (1989), De Soete and DeSarbo (1991), Wedel and DeSarbo (1992) and DeSarbo, Ramaswamy, and Chatterjee (1992) for the analysis of metric conjoint data or conjoint choice data over the more traditionally used two-stage procedure in which in the first stage coefficients are estimated for each individual, which are in the second stage grouped by some hierarchical or non-hierarchical clustering algorithm.

Describe a procedure for testing the hypotheses that:

(a) the profiles have zero-covariance within segments for the $S = 3$ segment solution of the DeSarbo, Wedel, Vriens, and Ramaswamy (1992) procedure for analyzing conjoint data;

(b) the preference judgments have equal variances within segments, given zero covariance, for this same procedure.

References

Addelman, S. 1962: Orthogonal main-effect plans for asymmetrical factorial experiment. *Technometrics*, 4, 21–46.

Aitkin, M., Anderson, D., and Hinde, J. 1981: Statistical modelling of data on teaching style (with discussion). *Journal of the Royal Statistical Society*, A 144, 419–461.

Aitkin, M. and Rubin, D. B. 1985: Estimation and hypothesis testing in finite mixture distributions. *Journal of the Royal Statistical Society*, B 47, 67–75.

Akaike, H. 1974: A new look at statistical model identification. *IEEE Transactions on Automatic Control*, AC-19, 716–23.

Andersen, E. B. 1982: Latent structure analysis: A survey. *Scandinavian Journal of Statistics*, 9, 1–12.

Anderson, E. W. and Sullivan, M. W. 1993: Customer satisfaction and retention across firms. *Marketing Science*, forthcoming.

Anderson, J. J. 1985: Normal mixtures and the number of clusters problem. *Computational Statistics Quarterly*, 2, 3–14.

Antoniak, C. E. 1974: Mixtures of dirichlet processes with applications to bayesian nonparametric problems. *Annals of Statistics*, 2, 1152–74.

Babakus, E. and Boller, G. W. 1992: An empirical assessment of the SERVQUAL scale, *Journal of Business Research*, 24, 253–68.

Bartholomew, D. J. 1987: *Latent Variable Models and Factor Analysis*, New York: Oxford University Press.

Basford, K. E. and McLachlan, G. J. 1985: The mixture method of clustering applied to three-way data. *Journal of Classification* 2, 109–25.

Baum, L. E. and Eagon, J. A. 1967: An inequality with applications to statistical estimation for probabilistic Markov processes and to a model for ecology. *Bulletin of the American Mathematical Society*, 73, 360–3.

Baum, L. E., Petrie, T., Soules, G., and Weiss, N. 1970: A maximization technique occurring in the statistical analysis of probabilistic functions of Markov chains. *Annals of Mathematical Statistics*, 41, 164–71.

Ben-Akiva, M. and Lerman, S. R. 1985: *Discrete Choice Analysis*, London: The MIT Press.

Besag, J. E. 1986: On the statistical analysis of dirty pictures (with discussion). *Journal of the Royal Statistical Society*, B, 48, 259–302.

Bhattacharya, C. G. 1967: A simple method for resolution of a distribution into its Gaussian components. *Biometrics*, 23, 115–35.

Boyles, R. A. 1983: On the convergence of the EM algorithm. *Journal of the Royal Statistical Society*, B 45, 47–50.

Bozdogan, H. 1987: Model selection and Akaike's information criterion (AIC): The general theory and its analytical extensions. *Psychometrika*, 52, 345–70.

Bozdogan, H. and Sclove, S. L. 1984: Multi-sample cluster analysis using Akaike's information criterion. *Annals of the Institute of Statistics and Mathematics*, 36, 163–80.

Bucklin, R. E. and Gupta, S. 1992: Brand choice, purchase incidence and segmentation: An integrated approach. *Journal of Marketing Research*, 29 (May), 201–16.

Bucklin, R. E., Gupta, S., and Siddarth, S. 1991: Segmenting purchase quantity behavior: A poisson regression mixture model. Working paper, University of California, Los Angeles, USA.

Bye, B. V. and Schechter, E. S. 1986: A latent Markov model approach to the estimation of response errors in multivariate panel data. *Journal of the American Statistical Association*, 81, 375–80.

Carman, J. M. 1990: Consumer perceptions of service quality: An assessment of the SERVQUAL dimensions, *Journal of Retailing*, 66, 1, 33–55.

Cassie, R. M. 1954: Some uses of probability paper for the graphical analysis of polymodel frequency distributions. *Australian Journal of Marine and Freshwater Research*, 5, 513–22.

Charlier, C. V. L. and Wicksel, S. D. 1924: On the dissection of frequency functions. *Arkiv für Mathematik, Astronomi och Fysik*, Bd 18, 6.

Clogg, C. C. 1979: Some latent structure models for the analysis of Likert-type data. *Social Science Research*, 8, 287–301.

Clogg, C. C. 1981: New developments in latent structure analysis. In D. J. Jackson and E. F. Borgatta (eds), *Factor Analysis and Measurement in Sociological Research*, London: Sage, 215–46.

Clogg, C. C. and Goodman, L. A. 1986: Latent structure analysis of a set of multidimensional contingency tables. *Journal of the American Statistical Association*, 79, 762–71.

Clogg, C. C. and Goodman, L. A. 1986: On scaling models applied to data from several groups. *Psychometrika*, 51, 123–35.

Coslett, S. R. and Lee, L. F. 1985: Serial correlation in discrete variable models. *Journal of Econometrics*, 27, 79–97.

Cronin, J. J. and S. A. Taylor 1992: Measuring service quality: A reexamination and extension, *Journal of Marketing*, 56, 55–68.

Croon, M. 1989: Latent class models for the analysis of rankings. In De Soete, G., Feger, H., and Klauer, K. C., *New Developments in Psychological Choice Modelling*, Amsterdam: Elsevier, North-Holland.

Day, N. E. 1969: Estimating the components of a mixture of two normal distributions. *Biometrika*, 56, 463–74.

Dayton, C. M. and MacReady, G. B. 1988: Concomitant variable latent class models. *Journal of the American Statistical Association*, 83, 173–8.

De Soete, G. and DeSarbo, W. S. 1991: A latent class probit model for analyzing pick and/N data. *Journal of Classification*, 8, 45–63.

Dempster, A. P., Laird, N. M. and Rubin, R. B. 1977: Maximum likelihood from incomplete data via the EM-algorithm. *Journal of the Royal Statistical Society*, B39, 1–38.

DeSarbo, W. S. and Cron, W. L. 1988: A maximum likelihood methodology for clusterwise linear regression. *Journal of Classification* 5: 249–82.

DeSarbo, W. S., Huff, L., Rolandelli, M., and Choi, J. 1993: On the measurement of perceived service quality: A conjoint analysis approach. In R. Rust and R. Oliver (eds), *Handbook of Service Quality*, Sage Press, forthcoming.

DeSarbo, W. S., Manrai, A., and Manrai, L. 1993: Latent class multidimensional scaling: a review of recent developments in the marketing and psychometric literature. In R. Bagozzi, (ed.), *Handbook of Marketing Research*, forthcoming.

DeSarbo, W. S., Oliver, R. L., and Rangaswamy 1990: A simulated annealing methodology for cluster wise linear regression. *Psychometrika*, 54, 707–36.

DeSarbo, W. S., Ramaswamy, V., and Chatterjee, R. 1992: Latent class multivariate conjoint analysis with constant sum data. *Working Paper, University of Michigan*, Ann Arbor, MI.

DeSarbo, W. S., Ramaswamy, V., Reibstein, D. J., and Robinson, W. T. 1992: A latent pooling methodology for regression analysis with limited time series of cross sections: A PIMS data application. *Marketing Science*, forthcoming.

DeSarbo, W. S., Wedel, M., Vriens, M., and Ramaswamy, V. 1992: Latent class metric conjoint analysis. *Marketing letters*, 3, 3, 273–88.

Dillon, W. R., Kumar, A., and de Borrero, M. S. 1993: Capturing individual differences in paired comparisons: An extended BTL model incorporating descriptor variables. *Journal of Marketing Research*, 30, 42–51.

Dillon, W. R. and Kumar, A. 1994: Latent structure and other mixture models in marketing: an integrative survey and overview. In R. P. Bagozzi (ed.), *Advanced Methods of Marketing Research*, Oxford: Blackwell.

Engel, C. and Hamilton, J. D. 1990: Long swings in the dollar: Are they in the data and do markets know it? *American Economic Review*, 80, 689–13.

Everitt, B. S. 1984: *An Introduction to Latent Variable Models*, London: Chapman & Hall.

Everitt, B. S. and Hand, D. J. 1981: *Finite Mixture Distributions*, London: Chapman & Hall.

Follmann, D. A. and Lambert, D. 1989: Generalizing logistic regression by non-parametric mixing. *Journal of the American Statistical Association*, 84, 295–300.

Formann, A. K. 1982: Linear logistic latent class analysis. *Biometrical Journal*, 20, 123–6.

Formann, A. K. 1984: *Die Latent-Class-Analyse*, Weinheim: Beltz.

Formann, A. K. 1985: Constrained latent class models: theory and applications. *British Journal of Mathematical and Statistical Psychology*, 38, 87–111.

Formann, A. K. 1992: Linear logistic latent class analysis for polytomous data. *Journal of the American Statistical Association*, 87, 476–86.

Fowlkes, E. B. 1979: Some methods for studying mixtures of two normal (lognormal) distributions. *Journal of the American Statistical Association*, 74, 561–75.

Fryer, I. G. and Robertson, C. A. 1972: A comparison of some methods for estimating mixed normal distributions. *Biometrika*, 59, 639–48.

Geman, S. and Geman, D. 1984: Markov random field image models and their applications to computational vision. *Proceedings of the International Congress on Mathematics*, American Mathematical Society.

Ghosh, J. M. and Sen, P. K. 1985: On the asymptotic performance of the log-likelihood ratio statistic for the mixture model and related results. *Proceedings of the Berkely Conference, Neyman and Kiefer, II*, Wadsworth, Montery, 789–806.

Goldfeld, S. M. and Quandt, R. E. 1973: A Markov model for switching regressions. *Journal of Econometrics*, 1, 3–16.

Goldfeld, S. M. and Quandt, R. E. 1976: *Studies in Nonlinear Estimation*, Cambridge, MA: Ballinger.

Goodman, L. A. 1974a: Exploratory latent structure analysis using both identifiable and unidentifiable models. *Biometrika*, 61, 215–31.

Goodman, L. A. 1974b: The analysis of systems of qualitative variables when some variables are unobservable. Part I: A modified latent structure approach. *American Journal of Sociology*, 79, 1179–1259.

Goodman, L. A. 1978: *Analyzing Qualitative/Categorical Data: Log-linear and Latent Structure Analysis*, Cambridge: Abt books.

Goodman, L. A. 1979: On the estimation of parameters in latent structure analysis. *Psychometrika*, 44 (1), 123–8.

Green, P. E. and Rao, V. R. 1971: Conjoint measurement for quantifying judgmental data, *Journal of Marketing Research*, 8, 355–63.
Gupta, A. K. and Miyawaki, T. 1978: On a uniform mixture model. *Biometrics*, 20, 631–7.
Gupta, S. and Chintagunta, P. K. 1992: On using demographic variables to determine segment membership in logit mixture models. Working paper, Cornell University, Ithaca, NY, USA.
Haberman, S. J. 1977: Maximum likelihood estimates in exponential response models. *Annals of Statistics*, 5, 815–41.
Haberman, S. J. 1979: *Analysis of Qualitative Data: Vol. 2. New Developments*, New York: Academic Press.
Haertel 1984: An application of latent class models to assessment data. *Applied Psychological Measurement*, 8, 333–46.
Hagenaars, J. A. 1990: *Categorical Longitudinal Data*, London: Sage.
Hamilton, J. D. 1989: A new approach to the economic analysis of nonstationary time series and the business cycle. *Econometrika*, 57, 357–84.
Hamilton, J. D. 1990: Analysis of time series subject to changes in regime. *Journal of Econometrics*, 45, 39–70.
Hamilton, J. D. 1991: A Quasi-Bayesian approach to estimating parameters for mixtures of normal distributions. *Journal of Business and Economic Statistics*, 9, 27–39.
Harding, I. P. 1948: The use of probability paper for the graphical analysis of polymodel frequency distributions. *Journal of the Marine Biological Association*, UK 28, 141–53.
Harris, C. M. 1983: On mixtures of geometric and negative binomial distributions. *Communications in Statistics*, A, 12, 987–1007.
Hasselblad, V. 1966: Estimation of parameters for a mixture of normal distributions. *Technometrics*, 8, 431–44.
Hasselblad, V. 1969: Estimation of finite mixtures of distributions from the exponential family. *Journal of the American Statistical Association*, 64, 1459–71.
Helsen, K., Jedidi, K., and DeSarbo, W. S. 1992: A new approach to country segmentation utilizing multinational diffusion patterns. Working paper, University of Michigan, Ann Arbor, MI.
Henna, J. 1985: On estimating the number of constituents of a finite mixture of continuous distributions. *Annals of the Institute of Statistics and Mathemathics*, 37, 235–40.
Hosmer, D. W. 1974: Maximum likelihood estimates of the parameters of a mixture of two regression lines. *Communications in Statistics*, 3, 995–1006.
John, S. 1970a: On analyzing mixed samples. *Journal of the American Statistical Association*, 65, 755–60.
John, S. 1970b: On identifying the population of origin of each observation in a mixture of observations of two Gamma populations. *Technometrics*, 12, 565–8.
Kamakura, W. A. 1991: Estimating flexible distributions of ideal-points with external analysis of preference. *Psychometrika*, 56, 419–48.
Kamakura, W. A. 1992: A clusterwise multinomial logit model for multiple locally-independent choice sets. Working Paper, University of Pittsburgh, Pittsburgh, USA.
Kamakura, W. A. and Novak, T. P. 1992: Value-system segmentation: Exploring the meaning of LOV. *Journal of Consumer Research*, 19, 119–32.

Kamakura, W. A. and Russell, G. J. 1989: A probabilistic choice model for market segmentation and elasticity structure. *Journal of Marketing Research*, 26, 379–90.

Kamakura, W. A., Wedel, M., and Agrawal, J. 1992: Concomitant variable latent class models for the external analysis of choice data. Research Memorandum of the Institute of Economic Research, Faculty of Economics, nr. 486, University of Groningen, Groningen, Netherlands.

Langeheine, R. 1988: New developments in latent class theory. In *Latent Trait and Latent Class Models*, New York: Plenum Press, 77–108.

Langeheine, R. and Rost, J. 1988: *Latent Trait and Latent Class Models*, Plenum Press, New York.

Lazarsfeld, P. F. and Henry, N. W. 1968: *Latent Structure Analysis*, New York: Houghton Mifflin.

Levinson, S. E., Rabiner, L. R. and Sondhi, M. M. 1983: An introduction to the application of the theory of probabilistic functions of a Markov process to automatic speech recognition. *Bell Syst. Technical Journal*, 62, 1035–74.

Li, L. A. and Sedransk, N. 1988: Mixtures of distributions: a topological approach. *Annals of Statistics*, 16, 1623–34.

Lindgren, G. 1978: Markov regime models for mixed distributions and switching regressions. *Scandinavian Journal of Statistics*, 5, 81–91.

Lüxmann-Ellinghaus, U. 1987: On the identifiability of mixtures of infinitely divisible power series distributions. *Statistical and Probability Letters*, 5, 375–8.

Lwin, T. and Martin, P. J. 1989: Probits of mixtures. *Biometrics*, 45, 721–32.

McCullagh, P. and Nelder, J. A. 1989: *Generalized Linear Models*, New York: Chapman & Hall.

McFadden, D. 1973: Conditional logit analysis of qualitative choice behavior. In P. Zarembarka (ed.), *Frontiers of Econometrics*, New York: Academic Press, 105–42.

McHugh, R. B. 1956: Efficient estimation and local identification in latent class analysis. *Psychometrika*, 21, 331–47.

McHugh, R. B. 1958: Note on "Efficient estimation and local identification in latent class analysis." *Psychometrika*, 23, 273–4.

McLachlan, G. J. 1982: The classification and mixture maximum likelihood approaches to cluster analysis. In P. R. Krishnaiah and L. N. Kanal (eds), *Handbook of Statistics* (vol. 2), Amsterdam: North-Holland, 199–208.

McLachlan, G. J. 1987: On bootstrapping the likelihood ratio test statistic for the number of components in a normal mixture. *Applied Statistics*, 36, 318–24.

McLachlan, G. J. and Basford, K. E. 1988: *Mixture Models: Inference and Application to Clustering*, New York: Marcel Dekker.

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

Mandelbaum, J. and Harris, C. M. 1982: Parameter estimation under progressive censoring conditions for a finite mixture of Weibull distributions. *TIMS/studies in Management Sciences*, 19, 239–60.

Mooijaart, A. and Heijden, P. G. M. van der 1992: The EM algorithm for latent class analysis with constraints. *Psychometrika*, 57, 261–71.

Nelder, J. A. and Wedderbum, R. W. M. 1972: Generalized linear models. *Journal of the Royal Statistical Society*, A–135, 370–84.

Newcomb, S. 1886: A generalized theory of the combination of observations so as to obtain the best result. *American Journal of Mathematics*, 8, 343–66.

Oliver, R. and DeSarbo, W. S. 1988: Response determinants in satisfaction judgments, *Journal of Consumer Research*, 14, 112–31.

Parasuraman, A., Zeithaml, V. A. and Berry, L. L. 1985: A conceptual model of service quality and its implications for future research, *Journal of Marketing*, 49 (Fall 1985), 41–50.

Parasuraman, A., Zeithaml, V. A. and Berry, L. L. 1988: SERVQUAL: A multiple-item scale for measuring consumer perceptions of service quality, *Journal of Retailing*, 64(1) (Spring), 12–40.

Pearson, K. 1894: Contributions to the mathematical theory of evolution. *Philosophical Trans.*, A 185, 71–110.

Pickett, E. E. and Whiting, R. G. 1987: On the estimation of probabilistic functions of Markov chains. In *Lecture Notes in Economics and Mathematical Syst.*, No. 297, Berlin: Springer.

Pol, F. J. R. van de, and Langeheine, R. 1989: Mixed Markov latent class models. In: C. C. Clogg (ed.), *Sociological Methodology*, Oxford: Blackwell, 213–47.

Pol, F. J. R. van de, and Leeuw, J. de 1986: A latent Markov model to correct for measurement error. *Sociological Methods and Research*, 15, 118–41.

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

Poulsen, C. A. 1990: Mixed Markov and latent Markov modelling applied to brand choice behavior. *International Journal for Research in Marketing*, 7, 5–19.

Quandt, R. E. and Ramsey, J. B. 1978: Estimating mixtures of normal distributions and switching regressions. *Journal of the American Statistical Association*, 73, 730–8.

Quandt, R. E. 1972: A new approach to estimating switching regressions. *Journal of the American Statistical Association*, 67, 306–10.

Ramaswamy, V., Anderson, E. W., and DeSarbo, W. S. 1993: Clusterwise negative binomial regression for count data analysis. *Management Science*, forthcoming.

Ramaswamy, V., DeSarbo, W. S., Reibstein, D. J., and Robinson, W. T. 1993: An empirical pooling approach for estimating marketing mix elasticities. *Marketing Science*, forthcoming.

Redner, R. A. and Walker, H. F. 1984: Mixture densities, maximum likelihood and the EM algorithm. *SIAM Review*, 26, 195–239.

Rindskopf, D. 1983: A general framework for using latent class analysis to test hierarchical and Non-hierarchical learning models. *Psychometrika*, 48, 85–97.

Schwartz, G. 1978: Estimating the dimensions of a model. *Annals of Statistics*, 6, 461–4.

Sclove, S. L. 1977: Population mixture models and clustering algorithms. *Communications in Statistics*, A6, 417–34.

Sclove, S. L. 1987: Application of model-selection criteria to some problems in multivariate analysis. *Psychometrika*, 52, 333–43.

Späth, H. 1979: Algorithm 39: Clusterwise linear regression. *Computing*, 22, 367–73.

Teicher, H. 1961: Identifiability of mixtures. *Annals of Mathematical Statistics*, 32, 244–48.

Titterington, D. M. 1990: Some recent research in the analysis of mixture distributions. *Statistics*, 4, 619–41.

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

Wedel, M. and DeSarbo, W. S. 1992a: A latent class binomial logit methodology for the analysis of paired comparison choice data. Memorandum from Institute of Economic Research, nr. 467, Faculty of Economics, University of Groningen, Groningen, Netherlands.

Wedel, M. and DeSarbo, W. S. 1992b: A mixture likelihood approach for generalized linear models. Memorandum from Institute of Economic Research, nr 478, Faculty of Economics, University of Groningen,

Wedel, M., DeSarbo, W. S., Bult, J. R., and Ramaswamy, V. 1991: A latent class Poisson regression model for heterogeneous count data. Memorandum from Institute of Economic Research, nr. 470, Faculty of Economics, University of Groningen, Groningen, Netherlands.

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

Wedel, M. and Leeflang, P. S. H. (1992), A pricing decision model for Gabor Granger price experiments. Memorandum from Institute of Economic Research, nr. 506, Faculty of Economics, University of Groningen, Groningen, Netherlands.

Wiggins, L. M. 1973: *Panel Analysis*, Amsterdam: Elsevier.

Windham, M. P. and Cutler, A. 1991: Information ratios for validating cluster analyses. *Conference paper, presented at the 1991 Joint Meetings of the Classification and Psychometric Societies*, Rutgers University, New York.

Wolfe, J. H. 1970: Pattern clustering by multivariate mixture analysis. *Multivariate Behavioral Research*, 5, 329–50.

Wu, C. F. J. 1983: On the convergence properties of the EM algorithm. *Annals of Statistics*, 11, 95–103.

Yarmal-Vuarl, F. and Ataman, E. 1987: Noise histogram and cluster validity for Gaussian-mixtured data. *Pattern Recognition*, 20, 385–401.