

On Estimating Finite Mixtures of Multivariate Regression and Simultaneous Equation Models

Kamel Jedidi

*Graduate School of Business
Columbia University*

Venkatram Ramaswamy and Wayne S. DeSarbo

*School of Business Administration
The University of Michigan*

Michel Wedel

*Department of Business Administration and Management Science
University of Groningen*

We propose a maximum likelihood framework for estimating finite mixtures of multivariate regression and simultaneous equation models with multiple endogenous variables. The proposed "semi-parametric" approach posits that the sample of endogenous observations arises from a finite mixture of components (or latent-classes) of unknown proportions with multiple structural relations implied by the specified model for each latent-class. We devise an Expectation-Maximization algorithm in a maximum likelihood framework to simultaneously estimate the class proportions, the class-specific structural parameters, and posterior probabilities of membership of each observation into each latent-class. The appropriate number of classes can be chosen using various information-theoretic heuristics. A data set entailing cross-sectional observations for a diverse sample of businesses is used to illustrate the proposed approach.

Multivariate and simultaneous equation models have been widely discussed in the statistical literature and numerous applications abound in the physical and social sciences (e.g., Bagozzi, 1980; Bentler, 1986; Goldberger & Duncan, 1973; Hurd, 1972; Jöreskog, 1977; Koenker & Portnoy, 1990; Raj, 1980). Simultaneous equation models may be viewed as multivariate regression models entailing endogenous variables that express the simultaneity in structural relations among the multiple dependent variables. The classic simultaneous equation model can be expressed as:

$$\underline{B} \underline{y}_i + \underline{\Gamma} \underline{x}_i = \underline{\zeta}_i, \quad i = 1, \dots, N (\text{cross-sectional observations}) \quad (1)$$

where $\underline{B} (Q \times Q)$ and $\underline{\Gamma} (Q \times P)$ are coefficient matrices and $\underline{\zeta}_i' = (\zeta_{i1}, \dots, \zeta_{iQ})$ is a random vector of disturbances assumed to be identically and independently distributed drawings from a multivariate normal distribution, $MVN(0, \underline{\Psi})$. $\underline{y}_i (Q \times 1)$ and $\underline{x}_i (P \times 1)$ are the vectors of endogenous and exogenous variables, respectively.¹ When there is no simultaneity (i.e., \underline{B} is diagonal), we obtain the classic multivariate linear regression model, of which the seemingly unrelated regression model is a special case (Zellner, 1962). Invoking the standard assumptions that \underline{B} is nonsingular and \underline{x}_i is uncorrelated with $\underline{\zeta}_i$, it can then be shown that \underline{y}_i follows a $MVN(-\underline{B}^{-1} \underline{\Gamma} \underline{x}_i, \underline{B}^{-1} \underline{\Psi} (\underline{B}^{-1})')$, and its density function can be written as:

$$f(\underline{y}_i) = \frac{|\underline{B}|}{(2\pi)^{Q/2} |\underline{\Psi}|^{1/2}} \exp \left(-\frac{1}{2} \left(\underline{B} \underline{y}_i + \underline{\Gamma} \underline{x}_i \right)' \underline{\Psi}^{-1} \left(\underline{B} \underline{y}_i + \underline{\Gamma} \underline{x}_i \right) \right) \quad (2)$$

It is common to estimate a single set of parameters \underline{B} , $\underline{\Gamma}$, and $\underline{\Psi}$ for the entire sample. Doing so, however, assumes homogeneity in the structural relations across the cross-sectional observations and may be justified if one is only interested in aggregate-level estimates. However, it is well-known that ignoring *cross-sectional heterogeneity* can potentially induce significant bias in the parameter estimates. One option is to introduce cross-sectional characteristics as moderating variables into the model formulation, or form homogeneous subgroups for analysis a priori. However, this requires elaborate theory and knowledge and may be operationally cumbersome, particularly if there is a large number of candidate variables. Another

¹Some readers may be more familiar with the following formulation of the structural equation model: $\underline{y} = \underline{B}' \underline{y} + \underline{\Gamma}' \underline{x} + \underline{\zeta}$. Comparing this formulation with Equation 1, it can be seen that $\underline{B} = (\underline{I} - \underline{B}')$, and $\underline{\Gamma} = -\underline{\Gamma}'$, and hence, the two formulations are equivalent.

appealing option, therefore, is to specify a random distribution of parameter values to capture any unobserved heterogeneity. The problem here is that the resulting formulation can be intractable and difficult to estimate depending on the specific parametric assumptions employed. Instead, one can assume a discrete mixture distribution for parameters over the entire sample (McLachlan & Basford, 1988). This type of "semi-parametric" approach can provide good numerical approximations even if the underlying mixing distribution is continuous (e.g., Heckman & Singer, 1984; Laird, 1978). The point masses of such a distribution represent the components of a finite mixture and can be interpreted as "latent-classes" of cross-sectional observations (e.g., Dayton & MacReady, 1988; DeSarbo & Cron, 1988; Wedel, DeSarbo, Bult, & Ramaswamy, 1993).

Finite mixture models have received considerable attention in the statistics literature (see DeSarbo, Manrai, & Manrai, 1994; Everitt & Hand, 1981; Titterton, Smith, & Makov, 1985; Wedel & DeSarbo, 1994, 1995, for comprehensive surveys and extensive discussions). These models postulate that a sample of observations arises from a finite mixture of underlying populations (of unknown proportions) with a specific form of densities in each population. Examples of such models that have been extensively used include finite mixtures of normal (e.g., Hasselblad, 1966; McLachlan, 1982), exponential (e.g., Teicher, 1961), and Bernoulli (e.g., Clogg & Goodman, 1984; Dayton and MacReady, 1988) densities. Over the past two decades, "conditional" mixture models have found wide appeal in the physical and social sciences (McCullagh & Nelder, 1989). Conditional mixture models allow for the probabilistic classification of observations and the simultaneous estimation of regression models relating the moments of the dependent variable for each mixture component to specified covariates. For instance, Quandt (1972), Hosmer (1974), and Quandt and Ramsey (1978) examined finite mixtures of univariate normal densities in which the expectations of these densities were specified as linear functions of covariates. To our knowledge, however, the various developments in formulating and estimating conditional mixtures for generalized linear models have been restricted to models entailing a single dependent variable (DeSarbo & Cron, 1988; Wedel & DeSarbo, 1995), and hence, cannot accommodate the model formulation implied by Equations 1 and 2.

In this article, we present the development of a maximum likelihood framework for estimating finite mixtures of multivariate regression and simultaneous equation models entailing *multiple endogenous variables*. Given a sample of heterogeneous cross-sectional observations, this method estimates K latent-classes of cross-sections with associated point masses, the class-specific structural parameters (B_k, Γ_k), the covariance matrix of disturbances, Ψ_k , and the posterior probability of membership of each cross-sectional unit into each of the derived classes. The variation in the structural parameters across the K classes captures the unobserved cross-sectional heterogeneity in the data. Separate sets of class-specific parameters

are estimated along with the size of each class. The estimation is performed using a maximum likelihood framework. The number of classes that adequately fits the data can be determined using various heuristics. A data set entailing cross-sectional observations for a diverse sample of businesses is used to illustrate the proposed finite mixture approach for estimating a simultaneous equation regression model with two endogenous variables.

THE PROPOSED FINITE MIXTURE MODEL

We assume that y_i is distributed as a finite mixture of conditional multivariate normal densities, $f_{ik}(\bullet)$:

$$y_i \sim \sum_{k=1}^K \lambda_k f_{ik}(y_i | x_i, B_k, \Gamma_k, \Psi_k) = \sum_{k=1}^K \lambda_k \left[\frac{|B_k|}{(2\pi)^{Q/2} |\Psi_k|^{1/2}} \exp \left(-\frac{1}{2} \left(B_k y_i + \Gamma_k x_i \right)' \Psi_k^{-1} \left(B_k y_i + \Gamma_k x_i \right) \right) \right] \quad (3)$$

where:

- $k = 1, \dots, K$ latent classes;
- $B_k = ((\beta_{mk}))$, the $(Q \times Q)$ matrix of endogenous variables coefficients for latent class k ($m = 1, \dots, Q$);
- $\Gamma_k = ((\gamma_{mj}))$, the $(Q \times P)$ matrix of exogenous variables coefficients for latent class k ($m = 1, \dots, Q; j = 1, \dots, P$);
- Ψ_k = the $(Q \times Q)$ covariance matrix of the disturbance vector ζ_k for latent class k ;
- $\lambda = (\lambda_1, \dots, \lambda_K)$, a vector of the K mixing proportions of the finite mixture (of which $K - 1$ are independent) such that $\lambda_k > 0$ and $\sum_{k=1}^K \lambda_k = 1$.

Assuming, in addition, that the y_i vectors are independent, the likelihood function for the N vectors (y_1, \dots, y_N) is given by:

$$L = \prod_{i=1}^N \left[\sum_{k=1}^K \lambda_k \left(\frac{|B_k|}{2\pi^{Q/2} |\Psi_k|^{1/2}} \exp \left(-\frac{1}{2} \left(B_k y_i + \Gamma_k x_i \right)' \Psi_k^{-1} \left(B_k y_i + \Gamma_k x_i \right) \right) \right) \right] \quad (4)$$

The mixing proportions $\hat{\lambda}$ can be construed as prior probabilities of any cross-sectional observation belonging to the K latent classes. Note, the posterior probability of membership for cross-section i in class k (\hat{P}_{ik}) can be computed within any iteration using Bayes' theorem, conditional on the estimates of the class-specific parameters $\hat{\lambda}_k$, \hat{B}_k , $\hat{\Gamma}_k$, and $\hat{\Psi}_k$ via:

$$\hat{P}_{ik} = \frac{\hat{\lambda}_k f_{ik}(y_i | x_i, \hat{B}_k, \hat{\Gamma}_k, \hat{\Psi}_k)}{\sum_{k=1}^K \hat{\lambda}_k f_{ik}(y_i | x_i, \hat{B}_k, \hat{\Gamma}_k, \hat{\Psi}_k)} \quad (5)$$

There are two important issues concerning the identifiability of the proposed finite mixture model. The first pertains to the identification of the finite mixture, whereas the second relates to the identification of the specified model with multiple endogenous variables when class membership is known. Regarding the first issue, because mixtures of multivariate normal densities are typically identified (see McLachlan & Basford, 1988, p. 97; Titterton et al., 1985, p. 162; Yakowitz & Spragins, 1968, p. 211), we can always identify the mixing proportions and the mean vectors ($\underline{\mu}_k$) and covariance matrices ($\underline{\Sigma}_k$) of the observables (y_i), regardless of the conditioning pattern imposed by the postulated simultaneous equation model. Further, given that the mean vectors and covariance matrices are known and are sufficient statistics for the simultaneous equation models in each class, the standard identification theory for multiple samples applies (Sörbom, 1974). Hence, if the data for each class are distributed as multivariate normal distribution, no new identification theory beyond that of multisample simultaneous equation models is required to recover the mixing proportions and other parameters of a latent-class simultaneous equation model. The lack of identifiability due to invariance of the likelihood under interchanging of the labels of the latent-classes (Aitkin & Rubin, 1985) is not an issue here, and we follow the solution of McLachlan and Basford (1988) in reporting results for only one of the possible arrangements of the classes.

Concerning the second issue, the model as specified in Equation 3 is not identified if all elements in $\underline{\Gamma} = (\Gamma_1, \dots, \Gamma_K)$, $\underline{B} = (B_1, \dots, B_K)$, and $\underline{\Psi} = (\Psi_1, \dots, \Psi_K)$ are free. Identification in this context, therefore, requires placing restrictions on model parameters. The most common restrictions set some elements of $\underline{\Gamma}$, \underline{B} , and $\underline{\Psi}$ to zero or some other constant, whereas others entail the imposition of equality or inequality constraints on parameters (Jöreskog, 1977). Identification rules, such as the order and rank conditions developed in the context of linear simultaneous equation models (see Greene, 1993, p. 592), can be utilized in the present context as well to establish the identifiability of the model formulation within each latent class.

THE EXPECTATION-MAXIMIZATION (EM) ALGORITHM FOR ESTIMATION

The likelihood of finite mixture models can be maximized using optimization methods such as the Newton-Raphson or by using the EM algorithm (Dempster, Laird, & Rubin, 1977). Convergence is not ensured with the former method (Atkinson, 1989; McLachlan & Basford, 1988), although relatively few iterations are required for convergence. The EM algorithm is attractive because it can be programmed easily and convergence is ensured; Dempster et al. (1977) provided a proof based on Jensen's inequality that the EM algorithm gives monotone increasing values of the likelihood function. However, the EM algorithm typically requires relatively more iterations and may converge to local optima, although several procedures have been proposed to improve the rate of convergence (e.g., Jones & McLachlan, 1992; Louis, 1982; Meilijson, 1989). Although either optimization method can be utilized for finite mixture models (Everitt, 1984), the EM algorithm has apparently been the most popular (Titterton, 1990).

We devised an EM-based algorithm for maximizing the log likelihood function for the proposed finite mixture model for estimating $\underline{\Gamma}_k$, \underline{B}_k , $\underline{\Psi}_k$, and $\underline{\lambda}(k = 1, \dots, K)$, given $\underline{y} = (y_1, \dots, y_N)'$, $\underline{x} = (x_1, \dots, x_N)'$, a value of K , the constraints imposed on $\underline{\lambda}$ earlier, and $|\underline{\Psi}_k| > 0$. The latter condition is necessary because the likelihood function is unbounded when $\underline{\Psi}_k$ is singular, and hence consistent estimators are not possible. Upon estimation, the maximum likelihood estimates $\hat{\underline{B}}_k$, $\hat{\underline{\Gamma}}_k$ characterize the structural relations within each class, and the estimated posterior probabilities of membership, $\hat{\underline{P}}_k$, provide a probabilistic allocation of the N cross-sectional observations into the K classes. A discrete allocation can then be obtained by assigning each cross-sectional unit to the modal class (i.e., the class with the largest posterior probability).

To formulate the EM-based algorithm, we define a latent class indicator variable z_{ik} as follows (cf. De Soete & DeSarbo, 1991; Jedidi, Ramaswamy, & DeSarbo, 1993):

$$z_{ik} = \begin{cases} 1 & \text{iff observation } i \text{ belongs to latent class } k, \\ 0 & \text{otherwise.} \end{cases}$$

We also assume that, for a particular observation i , the nonobserved vector, $\underline{z}_i = (z_{i1}, \dots, z_{iK})'$, is identically and independently multinomially distributed with probabilities $\underline{\lambda}$, that is,

$$(z_i | \lambda) \sim \prod_{k=1}^K \lambda_k^{z_{ik}} \tag{6}$$

Hence, the distribution of y_i given z_i is:

$$(y_i | z_i) \sim \prod_{k=1}^K z_{ik} f_{ik}(y_i | x_i, \Gamma_k, B_k, \Psi_k) = \prod_{k=1}^K \left[f_{ik}(y_i | x_i, \Gamma_k, B_k, \Psi_k) \right]^{z_{ik}} \tag{7}$$

Thus, considering $Z = ((z_{ik}))$ as missing data, we can write the complete log likelihood function as:

$$\begin{aligned} \ln L_c(\underline{A}, \underline{\Psi}, \underline{\lambda} | \underline{Z}, \underline{Y}, \underline{X}) &= \sum_{i=1}^N \sum_{k=1}^K z_{ik} \ln(f_{ik}(y_i | x_i, A_k, \Psi_k)) + \sum_{i=1}^N \sum_{k=1}^K z_{ik} \ln(\lambda_k) \tag{8} \\ &= \text{constant} - \frac{1}{2} \left[\sum_{i=1}^N \sum_{k=1}^K z_{ik} (\ln |\Psi_k| - \ln |B_k|^2 + (B_k y_i + \Gamma_k x_i)' \Psi_k^{-1} (B_k y_i + \Gamma_k x_i)) \right] + \sum_{i=1}^N \sum_{k=1}^K z_{ik} \ln(\lambda_k) \\ &= \text{constant} - \frac{1}{2} \left[\sum_{k=1}^K N_k \left[\ln(|\Psi_k|) - \ln(|B_k|^2) + \text{tr}(\Psi_k^{-1} A_k M_k A_k') \right] \right] + \sum_{k=1}^K N_k \ln(\lambda_k), \end{aligned}$$

where

$$\begin{aligned} N_k &= \sum_{i=1}^N z_{ik} \\ A_k &= (B_k, \Gamma_k) \\ \underline{A} &= (A_1, \dots, A_k) \\ M_k &= \sum_{i=1}^N \frac{z_{ik}}{N_k} (w_i w_i') \\ w_i &= \begin{pmatrix} y_i \\ x_i \end{pmatrix} \\ \underline{\Psi} &= (\Psi_1, \dots, \Psi_k) \end{aligned}$$

Equation 8 is equivalent to the log of Equation 4 when Z is observed. Because Z is unobserved, the maximization of $\ln L_c$ with respect to \underline{A} , $\underline{\Psi}$, and $\underline{\lambda}$ entails a two-step procedure. An E-step occurs where the expected value of z_{ik} , given provisional estimates for \underline{A} , $\underline{\Psi}$, and $\underline{\lambda}$, is substituted for z_{ik} . An M-step occurs next

where Equation 8 is maximized with respect to \underline{A} , $\underline{\Psi}$, and $\underline{\lambda}$, conditional on the new provisional estimates of z_{ik} , $i = 1, \dots, N$ and $k = 1, \dots, K$. These E- and M-steps are successively applied until convergence.

Using Bayes' rule, it can be easily shown that:

$$E\left(z_{ik} | y_i, \hat{\underline{A}}, \hat{\underline{\Psi}}, \hat{\underline{\lambda}}\right) = \frac{\hat{\lambda}_k f_{ik}(y_i | x_i, \hat{\underline{A}}_k, \hat{\underline{\Psi}}_k)}{\sum_{j=1}^K \hat{\lambda}_j f_{ij}(y_i | x_i, \hat{\underline{A}}_j, \hat{\underline{\Psi}}_j)} = \hat{P}_{ik} \quad (9)$$

Thus, in the E-step, we substitute \hat{P}_{ik} for z_{ik} in Equation 8.

In the M-step, we maximize Equation 8 with regard to \underline{A} , $\underline{\Psi}$, and $\underline{\lambda}$ subject to the restriction $\hat{\lambda}_k > 0$ and $\sum_k \hat{\lambda}_k = 1$, conditional on the new provisional estimates of z_{ik} . To estimate $\hat{\underline{\lambda}}$, it suffices to maximize the augmented function:

$$\sum_{k=1}^K \hat{N}_k \ln(\lambda_k) - \theta \left(\sum_{k=1}^K \lambda_k - 1 \right) \quad (10)$$

where θ denotes a Lagrangian multiplier and $\hat{N}_k = \sum_{i=1}^N \hat{P}_{ik} / N$. By differentiating Equation 10 with regard to θ and λ_k , and setting the derivatives equal to zero, we can show that:

$$\hat{\lambda}_k = \frac{\hat{N}_k}{N} \quad (11)$$

Given $\hat{\underline{P}} = ((\hat{P}_{ik}))$ and $\hat{\underline{\lambda}}$, the maximization of Equation 8 with regard to \underline{A} and $\underline{\Psi}$ can be converted into K minimization problems of the form:

$$\text{Min}_{\underline{A}_k, \underline{\Psi}_k} F_k = \frac{\hat{N}_k}{2} \left[\ln(|\underline{\Psi}_k|) - \ln(|\underline{B}_k|^2) + \text{tr}(\underline{\Psi}_k^{-1} \underline{A}_k \underline{M}_k \underline{A}_k') \right] \quad (12)$$

Here, we utilize the conjugate gradient method with automatic restarts (Powell, 1977) in the minimization of Equation 12. This method makes use of the first derivatives of F_k with regard to the free parameters in \underline{B}_k , $\underline{\Gamma}_k$, and $\underline{\Psi}_k$ in finding the minimum by an iterative search process. These partial derivatives are:

$$\frac{\partial F_k}{\partial B_k} = \hat{N}_k \left[\Psi_k^{-1} \left[B_k M_{yy}^t + \Gamma_k M_{xy}^t \right] - B_k^{-1} \right] \tag{13}$$

$$\frac{\partial F_k}{\partial \Gamma_k} = \hat{N}_k \left[\Psi_k^{-1} \left[B_k M_{yx}^t + \Gamma_k M_{xx}^t \right] \right] \tag{14}$$

$$\frac{\partial F_k}{\partial \Psi_k} = \hat{N}_k \left[\Psi_k^{-1} - \Psi_k^{-1} A_k M_k A_k' \Psi_k^{-1} \right] - \frac{\hat{N}_k}{2} \left[\text{Diag} \left(\Psi_k^{-1} - \Psi_k^{-1} A_k M_k A_k' \Psi_k^{-1} \right) \right] \tag{15}$$

where:

$$M_k = \begin{bmatrix} M_{yy}^t & M_{yx}^t \\ M_{xy}^t & M_{xx}^t \end{bmatrix} \tag{16}$$

The new estimates \hat{A} , $\hat{\Psi}$, and $\hat{\lambda}$ serve as the new provisional estimates for the next E-step iteration of the EM algorithm where the estimated posterior probabilities, \hat{P}_k , in Equation 9 are updated. The E-step and the M-step are alternated until no further improvement in the objective function in Equation 8 is possible. Hence, although convergence to at least a locally optimum solution is guaranteed, different starting values of the parameters must be used to investigate the potential occurrence of local optima.

The estimation algorithm must be executed for varying numbers of classes to determine the appropriate number of classes (K), because the actual number of classes is rarely known in practice. To test the null hypothesis (H_0) of K classes against the alternative hypothesis (H_1) of $(K + 1)$ classes, the standard likelihood ratio test statistic is inappropriate because it is not (asymptotically) a full rank quadratic form under H_0 (Aitkin & Rubin, 1985; Li & Sedransk, 1988). This problem is similar to counting the modes of a density estimate where one cannot determine a confidence interval for the number of modes (Donoho, 1988); rather, only a lower bound can be obtained.² For mixture models, Monte Carlo procedures have been utilized for testing the number of classes (Aitkin, Anderson, & Hinde, 1981; De Soete & DeSarbo, 1991; McLachlan, 1987) by comparing the likelihood ratio statistic from the real data with a distribution of that statistic obtained from a number of data sets generated under H_0 . Such procedures, however, are computationally cumbersome (cf. McLachlan & Basford, 1988), and the observed rejection rates do not quite conform to the intended levels under the null hypothesis (Titterton, 1990).

²We thank an anonymous reviewer for this point.

Another set of techniques for testing the number of components are those based on information criteria in which a penalty, proportional to the number of parameters estimated, is imposed on the maximized log-likelihood. These include, in increasing order of penalty magnitude, Akaike's Information Criterion (AIC; Akaike, 1974), Schwarz's (1978) Bayesian Information Criterion (BIC), and Bozdogan's (1987) Consistent AIC (CAIC). Although these criteria (particularly the AIC) rely on the same asymptotic properties as the likelihood ratio test (Sclove, 1987), they are often employed in practice as heuristics for selecting the number of mixture components (Titterton, 1990). Bozdogan (1987) noted that among information criteria, the AIC tends to lead to over-parameterized models. The BIC and CAIC are preferred because they take into account the sample size and tend to favor more parsimonious mixture models. These two heuristics are computed as :

$$\text{CAIC}_K = -2 \ln L + T_K (\ln N + 1) \quad (17)$$

$$\text{BIC}_K = -2 \ln L + T_K \ln N \quad (18)$$

where T_K is the number of free parameters that are estimated. Accordingly, one chooses the model for which the value of these heuristics is smallest.

Before interpreting this chosen solution, however, it is useful to examine the extent of fuzziness in class membership (when $K > 1$). An entropy-based measure (E_K) can be computed using the estimated posterior probabilities of membership (Ramaswamy, DeSarbo, Reibstein, & Robinson, 1993):

$$E_K = 1 - \left[\sum_i \sum_k -\hat{P}_{ik} \ln \hat{P}_{ik} \right] / N \ln K \quad (19)$$

E_K is a relative measure of fuzziness of the derived latent classes. It has a value of 0 when all the posterior probabilities are equal for each cross-section and a value of 1 when the derived classes are discrete. A value of E_K close to zero is cause for concern as it implies that the centroids of the K classes are not sufficiently separated. Because E_K is a nonlinear measure, even values around 0.50 to 0.60 for a given K -class solution can be indicative of sufficient separation, and hence, one may wish to also examine the distribution of the largest posterior membership probabilities across the cross-sectional units. Further, if using a wide range of (random) starting values does not produce well-separated classes, the assumption of discrete components may not be appropriate for the data at hand. Moreover, because the convergence to local optima can be exacerbated when the component densities are not well separated, it is important to examine whether the estimated latent classes are well separated before interpreting the chosen solution.

Finally, it is important to note that if any of the mixing proportions ($\hat{\lambda}_k$) are zero, the solution will be degenerate. This can occur if a large number of classes are

extracted. However, this problem is easily resolved by simply ignoring the null mixing proportions and reducing K accordingly. For the chosen K -class solution, we can compute the asymptotic variance-covariance matrix of the parameter estimates based on the asymptotic efficiency property of the maximum likelihood estimations. To examine the goodness of fit for the estimated model, we can compute the proportion of variance in each endogenous variable accounted for by the explanatory variables:

$$R_{iq}^2 = \text{Var}(\hat{y}_{iq}) / \text{Var}(y_{iq}) \quad (20)$$

where $\hat{y}_{iq} = \sum_{k=1}^K \hat{P}_{ik} (-\hat{B}_k^{-1} \hat{\Gamma}_k x_i)$ is the predicted value of the cross-sectional observation y_{iq} .

ILLUSTRATIVE APPLICATION

We present an illustrative application entailing the examination of price-quality relations in the marketplace.³ Although consumers may use price as a cue in the quality perception process (Geistfeld, 1988), Monroe and Dodds (1988) noted that consumers may judge the acceptability of price after assessing the quality of the product and suggest that the price-quality relation be modeled in a simultaneous fashion. For purposes of illustrating the proposed method, we focus on research involving perceived quality and price that has been conducted with data drawn from the Profit Impact of Marketing Strategy (PIMS) SPI4 database. The PIMS data are at the level of the *strategic business unit* (SBU), which is defined as an individual business that sells a distinct set of products to a served market and in competition with a well-defined set of competitors. A *served market* is defined as a segment of a total market wherein an identifiable group of customers have similar requirements for a product. The SPI4 database contains 4-year averages of data including several business-market characteristics and competitive strategies for over 3,000 participating SBUs (Buzzell & Gale, 1987).

Based on research studies using the PIMS database (e.g., Phillips, Chang, & Buzzell, 1983; Robinson & Fornell, 1985; Tellis & Fornell, 1988), we specify and estimate the simultaneous equation regression model summarized in Table 1. The operational definitions of the variables in Table 1 are summarized in Appendix A. The covariates can be classified into three categories (Tellis & Fornell, 1988): (a) strategic variables (relative marketing expenditures, new product introductions,

³We thank William Robinson for his suggestions regarding the illustrative application employed in this article.

TABLE 1
Specification of Regression Models

Independent Variable	Dependent Variable (Equation)	
	Relative Quality	Relative Price
Endogenous		
Relative perceived quality	—	β_{21}
Relative price	β_{12}	—
Exogenous		
Strategic:		
Relative marketing expenditures	γ_{11}	γ_{21}
Percentage from new products	γ_{12}	γ_{22}
Product patents	γ_{13}	—
Relative direct costs	—	γ_{24}
Market structural:		
Market pioneer	γ_{15}	γ_{25}
20-year pioneer	γ_{16}	γ_{26}
Follower	γ_{17}	γ_{27}
Market leadership	—	γ_{28}
Number of immediate customers	—	γ_{29}
Internal structural:		
Relative backward integration	γ_{110}	—
Value added/sales	γ_{111}	—

product patents, and relative direct costs), (b) market structure variables (order of market entry, market leadership, and number of immediate customers), and (c) internal structure variables (relative backward integration and value added/sales ratio). For the specification in Table 1, it can be shown that the rank and order conditions are satisfied indicating that the parameters of the model are identified.

We used a cross-sectional sample of 2,404 businesses in our illustrative application with a single 4-year average for each business. The sample entails businesses operating in different types of served markets, including consumer durable goods, nondurable household goods, and industrial products that span various traditional "industries" such as agriculture, mining, construction, tobacco, textiles, paper and allied products, printing, chemicals, primary and fabricated metals, plastics, mechanical and electrical machinery, transportation equipment, precision instruments, communications, and so on. Further, about one-fourth of the served markets are located outside of North America. Given the diverse cross-section of businesses and their operating environments, one might expect substantial cross-sectional *heterogeneity* in the structural relations, as captured by the parameters in Table 1.

As noted at the beginning of this article, a common practice is to obtain a single set of aggregate level estimates. These aggregate estimates (i.e., $K = 1$) for the present model are shown in Table 2. Scanning the estimates for the relative quality equation, note that relative price has a significant positive effect on relative quality.

TABLE 2
Parameter Estimates from Aggregate Model

<i>Variable</i>	<i>Relative Quality</i>	<i>Relative Price</i>
Relative perceived quality		0.341***
Relative price	0.315***	
Relative marketing expenditures	0.050**	0.095***
Percentage new products	0.025	0.007
Product patents	0.155***	
Relative direct costs		0.336***
Market pioneer	0.183***	-0.010
20-year pioneer	-0.087***	0.004
Follower	0.036	-0.015
Market leadership		0.131***
Number of immediate customers		0.028*
Relative backward integration	0.081***	
Value added/sales	0.099***	
Intercept	0.000	0.000

Note. Estimates are standardized.

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

Among the strategic exogenous variables, businesses that have higher relative marketing expenditures and benefit to a significant degree from patented products also tend to have higher quality. The order of entry effects indicate that market pioneers tend to have higher product quality, although this dissipates over time. The effects of the internal structure variables indicate that increased control over supply and internal operations enhances product quality. Scanning the estimates for the relative price equation in Table 2, note that relative quality has a significant positive impact on relative price. Among the strategic exogenous variables, relative direct costs has a relatively large impact on relative price; relative marketing expenditures also has a significant, albeit smaller, positive impact. Although the order of entry effects are insignificant, market leadership and number of immediate customers are significant and positive. Overall, there appears to be a positive relation between price and quality in the marketplace. The partial effects of price on quality (0.315), and quality on price (0.341), are almost equal in magnitude. This suggests a rather strong interdependence between price and quality, with each having an influence on the other.

Although these results are consistent with previous aggregate findings (Steenkamp, 1989; Tellis & Wernerfelt, 1987), such aggregate estimates can potentially mask cross-sectional variation in the structural relations if there is considerable heterogeneity in the data. For instance, Phillips et al. (1983) reported varying estimates of the partial effects of quality on price ranging from 0.190 to 0.480 depending on the type of business (the weighted average of these estimates is 0.355, which is close to our aggregate estimate of 0.341).

To accommodate cross-sectional heterogeneity and obtain disaggregate estimates of the structural relations, we applied the proposed finite mixture approach by varying the number of classes beyond the aggregate case of $K = 1$. The resulting ln-likelihood values, number of free parameters, BIC and CAIC values, as well as proportion of variance explained, are shown in Table 3, for $K = 1, \dots, 4$. Several starting values were used and the resulting parameter estimates were observed to converge around similar values. After inspection of these estimates and the log likelihood values, the "best" solution for each value of K was selected.

As is evident from Table 3, both the BIC and CAIC values are minimum for $K = 3$ classes, suggesting that three classes adequately describe the data. The last two columns of Table 3 give the proportion of variance explained (R^2) for the relative quality and relative price equations respectively. Note that although R^2 increases with the number of classes, there is only a marginal increase beyond $K = 3$, also providing additional support for the choice of the three-class solution. Although the entropy measure E_k is 0.625, the mean (across SBUs) of the largest (i.e., "modal") posterior probability for each SBU is 0.826. Further, about 82% of the SBUs have modal posterior probabilities greater than 0.70. The class-specific estimates (B_k, Γ_k) for the three-class solution are presented in Tables 4 and 5, for the relative quality and relative price equations respectively. The proportion of SBUs in these three classes (λ_k) are approximately 15%, 51%, and 34%, respectively.

For the sake of brevity, we concentrate primarily on interpreting the price-quality effects that are of substantial theoretical interest. From Table 4, note that the partial effect of relative price on relative quality is much larger for Class 2 (0.506) and Class 3 (0.559), whereas it is much smaller for Class 1 (0.157). From Table 5, note that the partial effect of relative quality on relative price is relatively large for Class 2 (0.382) and Class 3 (0.421), whereas it is somewhat smaller for Class 1 (0.267). Carpenter (1987) observed that higher quality brands, other things equal, price above lower quality brands. Our findings corroborate this general observation, which implies that prices convey important information about quality. However,

TABLE 3
Statistical Criteria for Choosing the Number of Classes

<i>Number of Classes (K)</i>	<i>Ln L</i>	<i>Q_K</i>	<i>BIC_K</i>	<i>CAIC_K</i>	<i>R² (Quality)</i>	<i>R² (Price)</i>
1	-6309.6	23	12798.3	12821.3	0.324	0.430
2	-5916.4	47	12198.7	12245.7	0.504	0.513
3	-5724.6	71	12001.9 ^a	12072.9 ^a	0.822	0.589
4	-5639.7	95	12018.9	12113.9	0.845	0.663

Note. BIC = Bayesian Information Criterion; CAIC = Consistent Akaike's Information Criterion.

^aMinimum BIC and CAIC.

TABLE 4
Parameter Estimates From Finite Mixture Model
for Relative Perceived Quality Equation

Variable	Class 1 (15%)	Class 2 (51%)	Class 3 (34%)
Relative price	0.157**	0.506**	0.559**
Relative marketing expenditures	0.006	0.051**	-0.010
Percentage new products	0.014	0.011	-0.018
Product patents	0.041*	0.072**	0.081**
Market pioneer	0.108**	0.222**	0.003
20-year pioneer	-0.011	-0.070*	-0.006
Follower	0.069*	0.061*	-0.009
Relative backward integration	0.037*	0.084**	0.024
Value added/sales	0.014	0.054**	0.064**
Intercept	1.541**	-0.082	-0.465**

Note. Estimates are standardized.

* $p \leq 0.05$. ** $p \leq 0.01$.

TABLE 5
Parameter Estimates From Finite Mixture Model
for Relative Price Equation

Variable	Class 1 (15%)	Class 2 (51%)	Class 3 (34%)
Relative perceived quality	0.267***	0.382***	0.421**
Relative marketing expenditures	0.094**	0.22*	0.113**
Percentage new products	0.073	-0.022*	0.013
Relative direct costs	0.475***	0.068***	0.292***
Market pioneer	0.086	-0.075***	0.026
20-year pioneer	0.012	0.051**	-0.091
Follower	-0.055**	-0.017	0.008
Market leadership	0.044	0.046***	0.331***
Number of immediate customers	-0.028	0.020***	0.070**
Intercept	0.028	-0.193***	0.337***

Note. Estimates are standardized.

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

the interdependence between price and quality is much stronger in Classes 2 and 3, and is relatively weaker in Class 1.

The aggregate ($K = 1$) estimates in Table 2 mask this heterogeneity in the estimated price-quality relation. Moreover, although the partial effects of price on quality and quality on price are about equal in magnitude in the aggregate, the disaggregate results from our proposed method suggest a potential asymmetry in price-quality relations. For Class 1, the impact of price on perceived quality is smaller than the impact of perceived quality on price. On the other hand, in Classes 2 and 3, prices have a relatively larger impact on perceived quality.

To formally diagnose the composition of the derived classes, we analyzed the posterior probabilities of membership via the following model (see Ramaswamy et al., 1993):

$$R_{ik} = \sum_r Z_{ir} \delta_{rk} + v_{ik} \quad (21)$$

where $R_{ik} = \ln(\hat{P}_{ik} / \hat{P}_i)$, and $\hat{P}_i = (\prod_r \hat{P}_{ir})^{1/K}$ is the geometric mean of the posterior probabilities, Z_{ir} is the value of the explanatory variable r for SBU i , δ_{rk} is the impact coefficient for variable r for pool k , and v_{ik} is a random normal disturbance. The explanatory variables we employed represented the characteristics of the product market and the scope of the business, which were expected to be associated with the posterior probabilities of membership for the derived classes, as discussed later. The specific operationalization of these explanatory variables are described in Appendix B. Note that we are able to conduct a formal analysis of the posterior probabilities because the PIMS database offers a rich source of background data on the businesses and their served markets.

The stage of the product life cycle (PLC) serves as a proxy for the level of information in the market because the proportion of informed consumers are typically higher in the later stages of the PLC, in part due to more marketing support that reduces search costs (Tellis & Wernerfelt, 1987). Tellis and Wernerfelt (1987) demonstrated analytically that the equilibrium correlation between price and quality is an increasing function of the level of information in the market. Hence, the correspondence between price and quality should increase over the duration of the PLC (Curry & Riesz, 1988). Because the strength of the price-quality relation should be weaker in Class 1, we expect businesses in the early stages of the PLC to have a higher likelihood of belonging to Class 1. Moreover, from Table 4, marketing expenditures has no significant impact on quality for Class 1, but is positively associated with price as seen in Table 5. This suggests that marketing expenditures tend to support price adjustments rather than changes in quality, which has been argued to be more likely in the early stages of the PLC (Schmalensee, 1982; Tellis & Fornell, 1988). This further suggests that businesses in the early stages of the PLC should have a higher likelihood of belonging to Class 1.

Other product-market characteristics, such as the type of goods sold, may also have a bearing on class membership. In industrial markets, buyers are more likely to be informed when their purchases entail raw or semi-finished materials, and particularly when auxiliary services such as customer education, installation, and repair are less important. In these situations, the level of information should be higher as products can be inspected on a regular basis (Nelson, 1970; Rangan, Moriarty, & Swartz, 1992), and such businesses should have a higher likelihood of belonging to Classes 2 and 3, which exhibit stronger price-quality relations. For

consumer markets, in the present context the prices reflect what immediate customers such as wholesalers and retailers pay for SBU products. Unlike the case of consumer durables that typically involve a large number of modest independent retail outlets as immediate customers, most of the consumer nondurables are sold to giant wholesalers, large retail chains, and institutional buying units who typically wield more buying power. Hence, it can be argued that the level of information in the consumer market, at the interface between the firm and channel intermediaries, is likely to be higher for nondurables than durables. Moreover, from Table 5, scanning the estimates for Class 3, market leadership appears to be an approach to commanding price premiums, along with increased marketing expenditures relative to competition. This phenomenon of pulling demand through the channel is characteristic of businesses selling nationally branded consumer nondurables (which comprise a majority of the consumer nondurable businesses in our sample), and serves to insulate them from price competition from lower priced, unadvertised products (e.g., Blattberg & Wisniewski, 1989; Bolton, 1989). Hence, we expect businesses selling consumer nondurables to be most likely to belong to Class 3 and least likely to belong to Class 2. Moreover, we expect these businesses to be in markets with a large number of competitors. In such situations, the supply of superior quality should enable businesses to command a premium (Tellis & Wernerfelt, 1987), which is corroborated by the relatively large partial effect of quality on price (0.421) for Class 3.

Apart from product-market characteristics, the scope of the business should also be associated with the derived classes. *Ceteris paribus*, if the business has a relatively narrow product or customer scope, the level of information in the SBU's served market is expected to be high. Consequently, we expect businesses that have a relatively narrow product scope or customer scope to have a smaller likelihood of belonging to Class 1.

The results of the regression analyses of the posterior probabilities of membership using the model specified in Equation 21 are given in Table 6. The impact coefficients have been standardized to facilitate relative comparisons. Comparing the likelihood ratio χ^2 for each pool with a critical χ^2 with 8 degrees of freedom, the results for all three pools are significant as shown in Table 6.

From Table 6, we find that businesses that are in the early stage of the PLC are indeed more likely to belong to Class 1. Also, Class 1 consists of markets where auxiliary services are more important whereas Class 3 consists of markets with a large number of competitors. However, although we expected the impact of the stage of the PLC to be weaker for Class 3, it is insignificant. Further, consistent with our expectations, businesses that sell consumer nondurables are more likely to belong to Class 3, whereas those that sell industrial raw materials are more likely to belong to Class 2. However, we find that consumer nondurables businesses are also likely to belong to Class 1, although the impact is smaller than that for Class 3. Finally, the scope of the business also has an impact. Businesses that have

TABLE 6
Analysis of Posterior Probabilities

<i>Variable</i>	<i>Class 1 (15%)</i>	<i>Class 2 (51%)</i>	<i>Class 3 (34%)</i>
Product/market characteristics			
Mature stage of PLC	-0.315*	0.266*	0.049
Consumer nondurables	0.144*	-0.348*	0.204*
Industrial raw materials	-0.308*	0.314*	-0.006
Importance of auxiliary services	0.188*	-0.220*	0.032
Number of competitors	-0.130*	0.030	0.101*
Business scope			
Broad product scope	0.282*	-0.125*	-0.157*
Broad customer scope	0.263*	-0.130*	-0.133*
Intercept	-2.346*	1.061*	1.285*
Likelihood ratio χ^2	207.270*	189.082*	86.278*

Note. PLC = product life cycle.

* $p \leq 0.01$.

relatively narrow product and customer scopes are indeed more likely to belong to Class 2 and Class 3, although the impact on Class 2 is somewhat larger.

Overall, it appears that a complex combination of both product-market characteristics and business scope are found to be associated with the composition of the derived classes. This underscores the difficulty of capturing unobserved cross-sectional heterogeneity a priori. In general, our empirical results seem to support the theoretical contention that the price-quality correlation varies with the level of information in the market. Tellis and Wernerfelt (1987) pointed out that this dependence is nonlinear and that a linear approximation would actually lead to weaker results. It should also be noted that the explanatory variables in our post hoc analysis serve merely as proxies for the information structure of product markets. Nevertheless, we are able to diagnose the derived classes and obtain some insights into the heterogeneity in price-quality relations across product markets.

CONCLUSION

Failure to recognize consumer heterogeneity in estimating regression models with multiple endogenous variables can potentially result in biased and misleading estimates at the aggregate level. The proposed approach affords a viable alternative for estimating discrete latent classes of cross-sections with similar structural relations for each class, along with the posterior probabilities of class membership for each cross-sectional unit. The structural relations for each cross-sectional unit can be viewed as a convex combination of the class-specific relations and the posterior probabilities of class membership for each cross-sectional unit. The

proposed approach can be extended to accommodate concomitant variables to provide more parsimonious representations of the data. For example, Dayton and MacReady (1988) discussed the treatment of such concomitant-variable latent-class models for categorical data.

Although the EM algorithm has good convergence properties (Boyles, 1983; Wu, 1983) and was well-behaved in the illustrative application, estimation problems can potentially arise in other situations with ill-conditioned data. Although it is possible to test equality of estimated parameters across the derived classes, the imposition of varying restrictions across the classes may be difficult to implement without prior knowledge of the composition of the classes. This underscores a pragmatic limitation of the proposed approach, in that varying restrictions between the classes can be difficult to specify prior to estimation, as the classes themselves are formed by the procedure. Despite such potential limitations, the proposed finite mixture approach can be gainfully utilized for estimating regression models with multiple endogenous variables, when faced with heterogeneous cross-sectional observations.

ACKNOWLEDGMENTS

Wayne S. DeSarbo now at the College of Business, Pennsylvania State University.
We thank the reviewers for their constructive comments and suggestions.

REFERENCES

- Aitkin, M., Anderson, D., & Hinde, J. (1981). Statistical modelling of data on teaching styles (with discussion). *Journal of the Royal Statistical Society, Series A, 144*, 419-461.
- Aitkin, M., & Rubin, D. B. (1985). Estimation and hypothesis testing in finite mixture distributions. *Journal of the Royal Statistical Society, Series B, 47*, 67-75.
- Akaike, H. (1974). A new look at statistical model identification. *IEEE Transactions on Automatic Control, 6*, 716-723.
- Atkinson, K. E. (1989). *An introduction to numerical analysis*. New York: Wiley.
- Bagozzi, R. P. (1980). *Causal models in marketing*. New York: Wiley.
- Bentler, P. M. (1986). Structural modeling and Psychometrika: An historical perspective on growth and achievements. *Psychometrika, 51*, 35-51.
- Blattberg, R. C., & Wisniewski, K. J. (1989). Price-induced patterns of competition. *Marketing Science, 8*, 291-309.
- Bolton, R. N. (1989). The relationship between market characteristics and promotional price elasticities. *Marketing Science, 8*, 153-169.
- Boyles, R. A. (1983). On the convergence of the EM algorithm. *Journal of the Royal Statistical Society, Series B, 45*, 47-50.
- Bozdogan, H. (1987). Model selection and Akaike's information criterion (AIC): The general theory and its analytical extensions. *Psychometrika, 52*, 345-370.
- Buzzell, R. D., & Gale, B. T. (1987). *The PIMS principles: Linking strategy to performance*. New York: Free Press.

- Carpenter, G. S. (1987). Modeling competitive marketing strategies: The impact of marketing-mix relationships and industry structure. *Marketing Science*, 6, 208–221.
- Clogg, C. C., & Goodman, L. A. (1984). Latent structure analysis of a set of multidimensional contingency tables. *Journal of the American Statistical Association*, 79, 762–771.
- Curry, D. J., & Riesz, P. C. (1988, January). Prices and price/quality relationships: A longitudinal analysis. *Journal of Marketing*, 52, 36–51.
- Dayton, C. M., & MacReady, G. B. (1988). Concomitant-variable latent-class models. *Journal of the American Statistical Association*, 83, 173–178.
- Dempster, A. P., Laird, N. M., & Rubin, D. B. (1977). Maximum likelihood from incomplete data via the EM algorithm. *Journal of the Royal Statistical Society, Series B*, 39, 1–38.
- DeSarbo, W. S., & Cron, W. L. (1988). A maximum likelihood methodology for clusterwise linear regression. *Journal of Classification*, 5, 249–282.
- DeSarbo, W. S., Manrai, A. K., & Manrai, L. A. (1994). Latent class multidimensional scaling: A review of recent developments in the marketing and psychometric literature. In R. P. Bagozzi (Ed.), *Advanced methods of marketing research* (pp. 190–222). Cambridge, MA: Blackwell.
- De Soete, G., & DeSarbo, W. S. (1991). A latent class probit model for analyzing pick any/n data. *Journal of Classification*, 8, 45–63.
- Donoho, D. L. (1988). One sided inference about functionals of a density. *The Annals of Statistics*, 16, 1390–1420.
- Everitt, B. S. (1984). Maximum likelihood estimation of the parameters in a mixture of two univariate normal distributions: A comparison of two different algorithms. *Statistician*, 33, 205–215.
- Everitt, B. S., & Hand, D. J. (1981). *Finite mixture distributions*. New York: Chapman and Hall.
- Geistfeld, L. V. (1988). The price–quality relationship: The evidence we have, the evidence we need. In E. S. Maynes et al. (Eds.), *The frontier of research in the consumer interest* (pp. 143–172). Columbia, MO: American Council on Consumer Interests.
- Goldberger, A. S., & Duncan, O. D. (1973). *Structural equation models in the social sciences*. New York: Academic.
- Greene, W. H. (1993). *Econometric analysis*. New York: Macmillan.
- Hasselblad, V. (1966). Estimation of parameters for a mixture of normal distributions. *Technometrics*, 8, 431–444.
- Heckman, J. J., & Singer, B. (1984). A method of minimizing the impact of distributional assumptions in econometric models for duration data. *Econometrica*, 52, 271–320.
- Hosmer, D. W. (1974). Maximum likelihood estimates of the parameters of a mixture of two regression lines. *Communications in Statistics*, 3, 995–1006.
- Hurd, M. D. (1972). Small-sample estimation of a structural equation with autocorrelated errors. *Journal of the American Statistical Association*, 67, 567–573.
- Jedidi, K., Ramaswamy, V., & DeSarbo, W. S. (1993). A maximum likelihood method for latent class regression involving a censored dependent variable. *Psychometrika*, 58, 375–394.
- Jones, P. N., & McLachlan, G. J. (1992). Improving the convergence rate of the EM algorithm for a mixture model fitted to grouped and truncated data. *Journal of Statistical Computation and Simulation*, 43, 31–44.
- Jöreskog, K. G. (1977). Structural equation models in the social sciences: Specification, estimation, and testing. In P. R. Krishnaiah (ed.), *Applications of statistics* (pp. 265–287). Amsterdam: North-Holland.
- Koenker, R., & Portnoy, S. (1990). M estimation of multivariate regressions. *Journal of the American Statistical Association*, 85, 1060–1068.
- Laird, N. (1978). Nonparametric maximum likelihood estimation of a mixing distribution. *Journal of the American Statistical Association*, 73, 805–811.
- Li, L. A., & Sedransk, N. (1988). Mixtures of distributions: A topological approach. *Annals of Statistics*, 16, 1623–1634.

- Louis, T. A. (1982). Finding the observed information matrix when using the EM algorithm. *Journal of the Royal Statistical Society, Series B*, 44, 226-233.
- McCullagh, P., & Nelder, J. A. (1989). *Generalized linear models*. New York: Chapman and Hall.
- McLachlan, G. J. (1982). The classification and mixture maximum likelihood approaches to cluster analysis. In P. R. Krishnaiah & L. N. Kanal (Eds.), *Handbook of statistics* (Vol. 2, pp. 199-208). Amsterdam: North-Holland.
- McLachlan, G. J. (1987). On bootstrapping the likelihood ratio test statistic for the number of components in a normal mixture. *Applied Statistics*, 36, 318-324.
- McLachlan, G. J., & Basford, K. E. (1988). *Mixture models: Inference and application to clustering*. New York: Marcel Dekker.
- Meilijson, J. (1989). A fast improvement of the EM algorithm on its own terms. *Journal of the Royal Statistical Society, Series B*, 51, 127-138.
- Monroe, K. B., & Dodds, W. B. (1988). A research program for establishing the validity of the price-quality relationship. *Journal of the Academy of Marketing Science*, 16, 151-168.
- Nelson, P. (1970). Information and consumer behavior. *Journal of Political Economy*, 78, 311-329.
- Phillips, L. W., Chang, D. R., & Buzzell, R. D. (1983). Product quality, cost position, and business performance: A test of some key hypotheses. *Journal of Marketing*, 47, 26-43.
- Powell, M. J. D. (1977). Restart procedures for the conjugate gradient method. *Mathematical Programming*, 12, 241-254.
- Quandt, R. E. (1972). A new approach to estimating switching regressions. *Journal of the American Statistical Association*, 67, 306-310.
- Quandt, R. E., & Ramsey, J. B. (1978). Estimating mixtures of normal regressions and switching regressions. *Journal of the American Statistical Association*, 73, 730-738.
- Raj, B. (1980). A Monte Carlo study of small-sample properties of simultaneous equation estimators with normal and non-normal disturbances. *Journal of the American Statistical Association*, 75, 221-229.
- Ramaswamy, V., DeSarbo, W. S., Reibstein, D. J., & Robinson, W. T. (1993, Winter). An empirical pooling approach for estimating marketing mix elasticities with PIMS data. *Marketing Science*, 12, 103-124.
- Rangan, V. K., Moriarty, R. T., & Swartz, G. S. (1992, October). Segmenting customers in mature industrial markets. *Journal of Marketing*, 56, 72-82.
- Robinson, W. T., & Fornell, C. (1985, August). Sources of market pioneer advantages in consumer goods industries. *Journal of Marketing Research*, 22, 305-317.
- Schmalensee, R. (1982, June). Product differentiation advantages of pioneering brands. *American Economic Review*, 54, 485-503.
- Schwarz, G. (1978). Estimating the dimension of a model. *Annals of Statistics*, 6, 461-464.
- Sclove, S. L. (1987). Application of model-selection criteria to some problems in multivariate analysis. *Psychometrika*, 52, 333-343.
- Sörbom, D. (1974). A general method for studying differences in factor means and factor structure between groups. *British Journal of Mathematical Psychology*, 27, 229-239.
- Steenkamp, J. B. (1989). *Product quality*. Assen/Maastricht, The Netherlands: Van Gorcum.
- Strategic Planning Institute. (1978). *Profit impact of marketing strategy data manual*. Cambridge, MA: Author.
- Teicher, H. (1961). Identifiability of mixtures. *Annals of Mathematical Statistics*, 31, 55-73.
- Tellis, G. J., & Fornell, C. (1988, February). The relationship between advertising and product quality over the product life cycle: A contingency theory. *Journal of Marketing Research*, 25, 64-71.
- Tellis, G. J., & Wernerfelt, B. (1987). Competitive price and quality under asymmetric information. *Marketing Science*, 6, 240-253.
- Titterton, D. M. (1990). Some recent research in the analysis of mixture distributions. *Statistics*, 4, 619-641.

- Titterton, D. M., Smith, A. F. M., & Makov, U. E. (1985). *Statistical analysis of finite mixture distributions*. New York: Wiley.
- Wedel, M., & DeSarbo, W. S. (1994). A review of recent developments in latent class regression models. In R. P. Bagozzi (Ed.), *Advanced methods of marketing research* (pp. 352–388). Cambridge, MA: Blackwell.
- Wedel, M., & DeSarbo, W. S. (1995). A mixture likelihood approach for generalized linear models. *Journal of Classification*, 12(1), 21–56.
- Wedel, M., DeSarbo, W. S., Bult, J. R., & Ramaswamy, V. (1993). A latent class poisson regression model for heterogeneous count data. *Journal of Applied Econometrics*, 8, 397–411.
- Wu, C. F. J. (1983). On the convergence properties of the EM algorithm. *The Annals of Statistics*, 11, 95–103.
- Yakowitz, S. J., & Spragins, J. D. (1968). On the identifiability of finite mixtures. *The Annals of Mathematical Statistics*, 39, 209–214.
- Zellner, A. (1962). An efficient method for estimating seemingly unrelated regressions and tests of aggregation bias. *Journal of the American Statistical Association*, 57, 500–509.

APPENDIX A:
Variables and Measures Used in the Empirical Application

<i>Variable</i>	<i>Definition</i>
Relative perceived quality	Difference between two percentage measures: proportion of SBU's sales volume due to goods of superior quality and proportion due to goods of inferior quality. Quality in both cases is defined in relation to the SBU's three major competitors and is judged from the consumer's perspective.
Relative price	Average price level of the SBU relative to the unweighted average price of the SBU's three major competitors.
Relative marketing expenditures	Average of annual advertising, promotion, and sales force expenditures relative to leading competitors.
Percentage new products	Percentage of the total sales of SBU accounted for by products introduced during the three preceding years.
Product patents	1 if the SBU benefits to a significant degree from patents and trade secrets; 0 otherwise.
Relative direct costs	The average level of the SBU's direct cost per unit of products and services relative to leading competitors. Includes costs of materials, production, and distribution.
Market pioneer	1 if the business is a market pioneer; 0 otherwise.
20-year pioneer	1 if the business is a market pioneer and has been in the market 20 years or more; 0 otherwise.
Follower	1 if the business is an early follower; 0 otherwise.
Market leadership	1 if the SBU has the largest market share; 0 otherwise.
Number of immediate customers	The number of immediate customers served by the business ranging from 1 (3 or fewer) to 8 (10,000 or more).
Relative backward integration	The degree of backward integration relative to leading competitors.
Value added/sales	The ratio of value added (sales minus purchases) to sales.

Note. For additional details, see Strategic Planning Institute's (1978) *Profit Impact of Marketing Strategy Data Manual*. SBU = strategic business unit.

APPENDIX B:
Variables and Measures Used to Explain the Latent Class Memberships

<i>Variable</i>	<i>Measure</i>
Product/market characteristics	
Mature stage of PLC	1 if business is in the mature stage of the PLC (demand growing at less than 10% annually in real terms); 0 otherwise.
Consumer nondurables	1 if the SBU sells household goods whose typical purchase amount is less than \$10; 0 otherwise.
Industrial raw materials	1 if the SBU sells raw materials to industrial OEMs; 0 otherwise.
Importance of auxiliary service	1 or 2 if auxiliary services (customer education, installation, and repair) provided are of some or great importance to end-users; 0 otherwise.
Number of competitors	The number of competing businesses, each with more than 1% of the served market.
Business scope	
Narrow product scope	1 if the breadth of product line is narrower than leading competitors; 0 otherwise.
Narrow customer scope	1 if the relative breadth of customer type is narrower than leading competitors or the relative number/sizes of customers are larger than leading competitors; 0 otherwise.

Note. PLC = product life cycle; SBU = strategic business unit; OEM = original equipment manufacturer.