Finite-Mixture Structural Equation Models for Response-Based Segmentation and Unobserved Heterogeneity

Kamel Jedidi • Harsharanjeet S. Jagpal • Wayne S. DeSarbo
Columbia University
Rutgers University
Pennsylvania State University

Abstract

Two endemic problems face researchers in the social sciences (e.g., Marketing, Economics, Psychology, and Finance): unobserved heterogeneity and measurement error in data. Structural equation modeling is a powerful tool for dealing with these difficulties using a simultaneous equation framework with unobserved constructs and manifest indicators which are error-prone. When estimating structural equation models, however, researchers frequently treat the data as if they were collected from a single population (Muthén 1989). This assumption of homogeneity is often unrealistic. For example, in multidimensional expectancy value models, consumers from different market segments can have different belief structures (Bagozzi 1982). Research in satisfaction suggests that consumer decision processes vary across segments (Day 1977).

This paper shows that aggregate analysis which ignores heterogeneity in structural equation models produces misleading results and that traditional fit statistics are not useful for detecting unobserved heterogeneity in the data. Furthermore, sequential analyses that first form groups using cluster analysis and then apply multigroup structural equation modeling are not satisfactory.

We develop a general finite mixture structural equation model that simultaneously treats heterogeneity and forms market segments in the context of a specified model structure where all the observed variables are measured with error. The model is considerably more general than cluster analysis, multigroup confirmatory factor analysis, and multigroup structural equation modeling. In particular, the model subsumes several specialized models including finite mixture simultaneous equation models, finite mixture confirmatory factor analysis, and finite mixture second-order factor analysis.

The finite mixture structural equation model should be of interest to academics in a wide range of disciplines (e.g., Consumer Behavior, Marketing, Economics, Finance, Psychology, and Sociology) where unobserved heterogeneity and measurement error are problematic. In addition, the model should be of interest to market researchers and product managers for two reasons. First, the model allows the manager to perform response-based segmentation using a consumer decision process model, while explicitly allowing for both measurement and structural error. Second, the model allows managers to detect unobserved moderating factors which account for heterogeneity. Once managers have identified the moderating factors, they can link segment membership to observable individual-level characteristics (e.g., socioeconomic and demographic variables) and improve marketing policy.

We applied the finite mixture structural equation model to a direct marketing study of customer satisfaction and estimated a large model with 8 unobserved constructs and 23 manifest indicators. The results show that there are three consumer segments that vary considerably in terms of the importance they attach to the various dimensions of satisfaction. In contrast, aggregate analysis is misleading because it incorrectly suggests that except for price all dimensions of satisfaction are significant for all consumers. Methodologically, the finite mixture model is robust; that is, the parameter estimates are stable under double cross-validation and the method can be used to test large models. Furthermore, the double cross-validation results show that the finite mixture model is superior to sequential data analysis strategies in terms of goodness-of-fit and interpretability.

We performed four simulation experiments to test the robustness of the algorithm using both recursive and nonrecursive model specifications. Specifically, we examined the robustness of different model selection criteria (e.g., CAIC, BIC, and GFI) in choosing the correct number of clusters for exactly identified and overidentified models assuming that the distributional form is correctly specified. We also examined the effect of distributional misspecification (i.e., departures from multivariate normality) on model performance. The results show that when the data are heterogeneous, the standard goodness-of-fit statistics for the aggregate model are not useful for detecting heterogeneity. Furthermore, parameter recovery is poor. For the finite mixture model, however, the BIC and CAIC criteria perform well in detecting heterogeneity and in identifying the true number of segments. In particular, parameter recovery for both the measurement and structural models is highly satisfactory. The finite mixture method is robust to distributional misspecification; in addition, the method significantly outperforms aggregate and sequential data analysis methods when the form of heterogeneity is misspecified (i.e., the true model has random coefficients).

Researchers and practitioners should only use the mixture methodology when substantive theory supports the structural equation model, *a priori* segmentation is infeasible, and theory suggests that the data are heterogeneous and belong to a finite number of unobserved groups. We expect these conditions to hold in many social science applications and, in particular, market segmentation studies.

Future research should focus on large-scale simulation studies to test the structural equation mixture model using a wide range of models and statistical distributions. Theoretical research should extend the model by allowing the mixing proportions to depend on prior information and/or subject-specific variables. Finally, in order to provide a fuller treatment of heterogeneity, we need to develop a general random coefficient structural equation model. Such a model is presently unavailable in the statistical and psychometric literatures.

(Structural Equation Models; Market Segmentation; Finite Mixture Models; Confirmatory Factor Analysis; Customer Satisfaction)

0732-2399/97/1601/0039$05.00
Copyright © 1997, Institute for Operations Research and the Management Sciences

Copyright © 2001 All Rights Reserved

1. Introduction

Structural equation models provide a powerful framework for estimating consumer decision process models and systems of simultaneous equations with measurement error. When estimating structural equation models, researchers frequently treat the data as if they were collected from a single population (Muthén 1989). This assumption of homogeneity is often unrealistic. For example, in multidimensional expectancy value models (Bagozzi 1982) consumers from different market segments can have different belief structures. Hence pooling the data across respondents is likely to produce misleading results. Several authors in the salesforce literature have cautioned against pooling data across salespeople when testing theories of salespeople's performance/satisfaction (e.g., Kohli 1989). Research in the satisfaction literature also suggests that consumer decision processes vary across segments (Day 1977).

To illustrate the problems stemming from failure to treat heterogeneity in the context of structural equation models, consider the model shown in Figure 1 where the unobserved Affect (η) for a new food product depends on two perceptual dimensions: Sweetness (ξ_1) and Richness (ξ_2). Suppose there are two unobserved benefit segments of equal size, indexed by $g = 1$ and $g = 2$, respectively. Segment 1 is "pleasure-seeking" whereas segment 2 is "health-conscious." Affect, Sweetness, and Richness are each measured by two indicators. The values of the factor loadings, structural parameters (i.e., the importance weights of the perceptual dimensions), and the factor covariances for both groups are shown in Figure 1. The variances of the measurement errors (i.e., the δ's and ϵ's) and the structural errors (i.e., the ζ's) are all equal to 0.5 and the variances of the exogenous factors are all equal to unity for both groups. The expected values of the exogenous factors ξ_1 and ξ_2 equal zero for group 1 (i.e., $E(\xi_i^1) = 0, i = 1, 2$) and the corresponding values for group 2 are 0.6 (i.e., $E(\xi_i^2) = 0.6, i = 1, 2$). Note that both Sweetness and Richness have positive impacts on Affect in segment 1 ("pleasure-seeking") and negative effects on Affect in segment 2 ("health-conscious").

Suppose we fail to recognize the heterogeneity and analyze the pooled covariance matrix.[1] Assuming a sample size of 500, we find that the fit statistics are perfect: $\chi_6^2 = 0$ ($p = 1.00$), goodness-of-fit index GFI = 1.00, and root mean residual RMR = 0, incorrectly implying that the data are homogeneous and the model is correct. The structural parameter estimates, however, are severely biased. Both estimates equal 0.054 and are statistically insignificant providing strong support for the erroneous theory that the perceptual dimensions Sweetness and Richness have no impact on Affect. Thus the aggregate structural equation model leads to meaningless results: in particular, the estimated parameters do not reflect the parameters of either segment even though the fit statistics using the pooled data are perfect. More seriously, the traditional fit statistics do not provide diagnostic information alerting the researcher to the presence of unaccounted heterogeneity in the model.

This paper develops a finite mixture structural equation model for detecting and treating unobserved heterogeneity. Substantively, the model allows the manager to perform response-based market segmentation using a consumer decision process model, while explicitly allowing for both measurement and structural error.[2] The models can range in complexity from reduced-form (i.e., single-equation) models with no measurement error to nonrecursive systems of simultaneous equations with measurement error. Methodologically, the model contributes to the consumer behavior literature by allowing researchers to detect unobservable discrete moderating factors which account for heterogeneity among consumers.[3] The model also contributes to the statistical literature by generalizing the multigroup structural equation model (Jöreskog 1971, Sörbom 1974) to the case where group membership is unknown.

Section 2 discusses alternative methods for treating heterogeneity in structural equation models. Section 3

[1] Let $\boldsymbol{\mu}_g$ and Σ_g, respectively, define the mean vectors and covariance matrices of the observable variables for group g. Then the pooled covariance matrix is

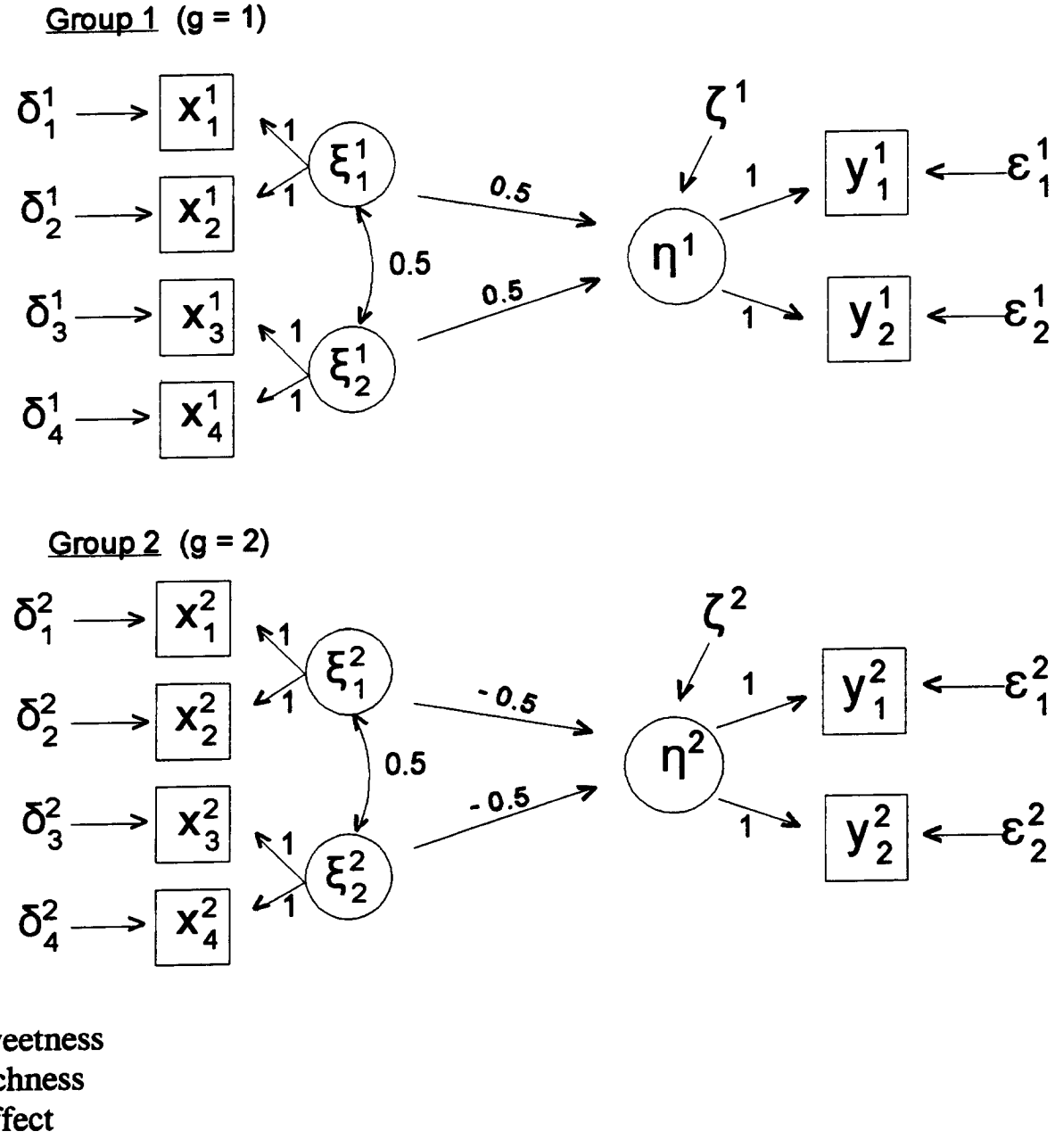
$$\Sigma_M = \sum_{g=1}^{2} 0.5\Sigma_g + 0.25(\boldsymbol{\mu}_1 - \boldsymbol{\mu}_2)(\boldsymbol{\mu}_1 - \boldsymbol{\mu}_2)'.$$

[2] The model differs from previous response-based segmentation models (e.g., Kamakura and Russell 1989 and Chintagunta, Jain, and Vilcassim 1991) because it allows for simultaneous equations and measurement error.

[3] We thank an anonymous referee for drawing our attention to this interpretation.

Copyright © 2001 All Rights Reserved

Figure 1 **An Example Illustrating the Effects of Unobserved Heterogeneity**



develops the general finite mixture structural equation model and shows how other models are special cases (e.g., finite mixture confirmatory factor analysis). Section 4 discusses model estimation and model selection criteria. Section 5 applies the finite mixture method to a commercial consumer satisfaction study. Section 6 discusses the results from simulations testing the robustness of the methodology. Finally, §7 summarizes the results, discusses limitations, and suggests future research directions.

Copyright © 2001 All Rights Reserved

2. Methods for Treating Heterogeneity in Structural Equation Models

Suppose the market researcher has formulated a structural equation model based on a well-developed theory. The researcher, however, suspects that the data belong to a finite number of groups / segments or, alternatively, that a discrete unobservable moderating factor accounts for heterogeneity.[4] The preferred data analysis strategy is first to form groups *a priori* and then use the standard multigroup structural equation methodology (Jöreskog 1971, Sörbom 1974). However, *a priori* segmentation based on demographic and psychographic variables may be infeasible or insufficient to explain differences in responses (Moore 1980).

Suppose the researcher cannot form segments *a priori* and does not know the number of groups or the moderating factors which account for heterogeneity. Two procedures are available. First, the researcher can perform cluster analysis on the indicator variables to form segments and then use multigroup structural equation modeling to estimate the model for each segment. This approach is problematic because "when clustering samples from a population, no cluster method is *a priori* believable without a statistical model" (see Aitkin, Anderson and Hinde 1981; cited in McLachlan and Basford 1988, pp. 2–3). Second, the researcher can use a variety of data-reduction methods (e.g., principal components or factor analysis) to filter the aggregate data by purging measurement error, form clusters (segments) using the reduced dimensions, and then perform multigroup structural equation modeling. This approach is logically flawed because the first step assumes homogeneity (i.e., one population) whereas the second asserts heterogeneity (i.e., multiple segments or populations). Empirical studies also show that this data-reduction strategy is not robust (Chang 1983). These difficulties are compounded when the data include measurement error (Baker 1974). For details on the efficiency of clustering in the presence of measurement error, see the survey papers by Milligan and Cooper (1987) and Bock (1996).

In addition to these statistical difficulties, current clustering methods are managerially restrictive: they do not allow the manager to perform response-based segmentation on the basis of a hypothesized model structure. This weakness is recognized in the literature. McLachlan and Basford (1988, p. 173) observe that, "Ideally one would like to perform a clustering of the entities *on the basis of all the information available*, assuming that differentiation between the groups is to be with respect to the total information." (Emphasis added.)

In sum, researchers can use the standard structural equation methodology when theory is sufficiently developed to allow them to form segments *a priori* or identify moderating factors which account for consumer heterogeneity. In many marketing applications, however, these strategies are not feasible. Hence we need a methodology for detecting and treating unobserved heterogeneity in structural equation models.

3. The Finite Mixture Structural Equation Model

We now propose a finite mixture structural equation model which deals with the problems caused by unobserved heterogeneity (see previous section) and generalizes the multigroup structural equation model (Jöreskog 1971, Sörbom 1974) to the case where group membership cannot be determined *a priori*.[5] Specifically, our method is a *model-based* approach for clustering using structured data. Hence, we can simultaneously form clusters (segments) and obtain segment-specific estimates for the measurement and structural parameters

[4] The true model could also consist of one population with random measurement and structural coefficients. Extant structural equation models cannot handle this form of heterogeneity except in trivial cases (i.e., single-equation models with no measurement error). Although we shall not develop a random coefficient structural equation model, our simulation studies will examine the robustness of the finite mixture model when the true structural equation model has random coefficients. We thank the Area Editor and an anonymous referee for suggesting this simulation study.

[5] Muthén (1989) proposed a MIMIC approach for handling certain forms of heterogeneity in structural equation models. Using this approach, the researcher can analyze the pooled data after including additional explanatory variables (e.g., demographics) to capture heterogeneity. The MIMIC approach, however, only allows for special forms of heterogeneity. Specifically, only the intercepts and the factor means can vary across consumers. That is, all consumers must have the same covariance matrices.

Copyright © 2001 All Rights Reserved

of a postulated model structure.[6] We can also use the method to test whether unobserved moderating factors account for heterogeneity in the structured data.

For convenience, we shall use the standard notation for multigroup structural equation models. Let g index membership in a (unknown) segment ($g = 1, \cdots G$), $\boldsymbol{\xi}^g$ denote an ($n \times 1$) vector of exogenous latent variables, and $\boldsymbol{\eta}^g$ denote an ($m \times 1$) vector of endogenous latent variables. Let $\mathbf{x}|g$ denote a ($q \times 1$) vector of observable indicator variables measuring $\boldsymbol{\xi}^g$ and $\mathbf{y}|g$ denote a ($p \times 1$) vector of observable indicator variables measuring $\boldsymbol{\eta}^g$. Let the measurement model be:

$$\mathbf{y}|g = \boldsymbol{\nu}_y^g + \boldsymbol{\Lambda}_y^g \boldsymbol{\eta}^g + \boldsymbol{\varepsilon}^g \tag{1}$$

$$\mathbf{x}|g = \boldsymbol{\nu}_x^g + \boldsymbol{\Lambda}_x^g \boldsymbol{\xi}^g + \boldsymbol{\delta}^g \tag{2}$$

where $\boldsymbol{\Lambda}_y^g(p \times m)$ and $\boldsymbol{\Lambda}_x^g(q \times n)$ are coefficient matrices (factor loadings), $\boldsymbol{\nu}_y^g(p \times 1)$, and $\boldsymbol{\nu}_x^g(q \times 1)$ are vectors of measurement intercept terms, and $\boldsymbol{\varepsilon}^g$ and $\boldsymbol{\delta}^g$ are vectors of measurement errors in $\mathbf{y}|g$ and $\mathbf{x}|g$, respectively. Let $E(\boldsymbol{\xi}^g) = \boldsymbol{\tau}_\xi^g$, $E[(\boldsymbol{\xi}^g - \boldsymbol{\tau}_\xi^g)(\boldsymbol{\xi}^g - \boldsymbol{\tau}_\xi^g)'] = \boldsymbol{\Phi}^g$, $E(\boldsymbol{\varepsilon}^g\boldsymbol{\varepsilon}^{g'}) = \boldsymbol{\Theta}_\varepsilon^g$, and $E(\boldsymbol{\delta}^g\boldsymbol{\delta}^{g'}) = \boldsymbol{\Theta}_\delta^g$ where $\boldsymbol{\Theta}_\varepsilon^g$ and $\boldsymbol{\Theta}_\delta^g$ are not necessarily diagonal. Assume that $E(\boldsymbol{\varepsilon}^g) = \mathbf{0}$, $E(\boldsymbol{\delta}^g) = \mathbf{0}$, and that the vectors of measurement errors are uncorrelated with the vector of latent variables.

The structural model defines the hypothesized theoretical links among the latent variables. Let the structural model be defined by:[7]

$$\mathbf{B}_g\boldsymbol{\eta}^g = \boldsymbol{\alpha}^g + \boldsymbol{\Gamma}^g\boldsymbol{\xi}^g + \boldsymbol{\zeta}^g, \tag{3}$$

where $\mathbf{B}_g$ is an ($m \times m$) matrix of structural parameters specifying the links among the endogenous latent variables, $\boldsymbol{\Gamma}^g$ is an ($m \times n$) coefficient matrix denoting the effect of $\boldsymbol{\xi}^g$ on $\boldsymbol{\eta}^g$, $\boldsymbol{\alpha}^g$ is an ($m \times 1$) vector of intercept terms, and $\boldsymbol{\zeta}^g$ is a random vector of disturbances (errors in equations). Let $E(\boldsymbol{\zeta}^g\boldsymbol{\zeta}^{g'}) = \boldsymbol{\psi}^g$. By assumption, $E(\boldsymbol{\zeta}^g) = \mathbf{0}$, $\boldsymbol{\zeta}^g$ is uncorrelated with $\boldsymbol{\xi}^g$, and $\mathbf{B}_g$ is nonsingular.

[6] Alternatively, one can use a two-step method first to form the groups without considering the structural model and then apply the multigroup structural equation methodology to the partitioned data. This procedure is statistically inefficient and may even be infeasible for large models.

[7] Strictly, we should use a superscript for g for the matrix $\mathbf{B}$. Instead, we have used a subscript to minimize notational clutter in our derivations, given the types of matrix operations performed on $\mathbf{B}_g$.

The finite mixture structural equation model defined by Equations (1) through (3) is very general and subsumes a variety of specialized models. Suppose the model is defined by Equation (1) or equivalently by Equation (2). The system now reduces to a finite mixture confirmatory factor model. Suppose all measures are error-free (i.e., $\boldsymbol{\nu}_y^g = \mathbf{0}$, $\boldsymbol{\nu}_x^g = \mathbf{0}$, $\boldsymbol{\Lambda}_x^g = \mathbf{I}$, $\boldsymbol{\Lambda}_y^g = \mathbf{I}$, $\boldsymbol{\Theta}_\delta^g = \mathbf{0}$, and $\boldsymbol{\Theta}_\varepsilon^g = \mathbf{0}$). The system reduces to Equation (3), a finite mixture simultaneous equation model. Suppose the model consists of Equations (1) and (3) only. The result is a finite mixture, second-order confirmatory factor model. Suppose the model is defined by the system of Equations (1), (2), and (3). This formulation allows for general types of heterogeneity in both the measurement *and* structural equation models (e.g., a model where consumers use different decision-process models and the factor mean structures differ across segments).

Before proceeding to model estimation, it is necessary to establish that the postulated finite mixture structural equation model is identified. Suppose the multigroup model for known groups is identified (see Sörbom 1974 for a discussion). Then the finite mixture of structural equation models is identified provided the data for the unknown groups follow multivariate normal distributions. This result follows from Definition 3.1.1 in Titterington, Smith, and Makov (1985, p. 36). See Appendix 1 for a proof.

4. Model Estimation and Model Selection

We now develop an empirical methodology for estimating the general finite mixture structural equation model (see Equations (1), (2), and (3)).

Model Estimation

Consider the general model defined by Equations (1), (2), and (3). Let $\boldsymbol{\Delta}|g = \begin{bmatrix}\mathbf{y}|g \\ \mathbf{x}|g\end{bmatrix}$ denote the joint vector of observable indicator variables conditional on membership in group g. Then its conditional mean vector $\boldsymbol{\mu}_g[(p + q) \times 1]$ is defined by:

$$\boldsymbol{\mu}_g = \begin{bmatrix} \boldsymbol{\nu}_y^g + \boldsymbol{\Lambda}_y^g\mathbf{B}_g^{-1}(\boldsymbol{\alpha}^g + \boldsymbol{\Gamma}^g\boldsymbol{\tau}_\xi^g) \\ \boldsymbol{\nu}_x^g + \boldsymbol{\Lambda}_x^g\boldsymbol{\tau}_\xi^g \end{bmatrix} \tag{4}$$

and its conditional covariance matrix $\boldsymbol{\Sigma}_g[(p + q) \times (p + q)]$ by:

Copyright © 2001 All Rights Reserved

$$\Sigma_g = \begin{bmatrix} \Lambda_y^g B_g^{-1}(\Gamma^g \Phi^g \Gamma^{g'} + \psi^g) B_g^{-1'} \Lambda_y^{g'} + \Theta_\varepsilon^g & \Lambda_y^g B_g^{-1} \Gamma^g \Phi^g \Lambda_x^{g'} \\ \Lambda_x^g \Phi^g \Gamma^{g'} B_g^{-1'} \Lambda_y^{g'} & \Lambda_x^g \Phi^g \Lambda_x^{g'} + \Theta_\delta^g \end{bmatrix}. \quad (5)$$

By assumption $\Delta | g$ has a conditional multivariate normal distribution. Hence the unconditional distribution of the observed vector $\Delta = [{}^{y}_{x}]$ is a finite mixture of these distributions. That is:

$$\Delta \sim \sum_{g=1}^{G} w_g f_g(\Delta | \mu_g, \Sigma_g), \quad (6)$$

where $\mathbf{w} = (w_1, \ldots, w_G)'$ is the vector of the G mixing proportions such that $w_g > 0$ and $\Sigma_{g=1}^{G} w_g = 1$, and $f(\cdot)$ is the conditional multivariate normal density function. The likelihood function for a sample $(\Delta_1, \ldots, \Delta_N)$ of $i = 1, \ldots, N$ randomly drawn observations from the mixture is then:

$$L = \prod_{i=1}^{N} \left[\sum_{g=1}^{G} w_g (2\pi)^{-(p+q)/2} |\Sigma_g|^{-1/2} \times \exp\{-\tfrac{1}{2}(\Delta_i - \mu_g)' \Sigma_g^{-1} (\Delta_i - \mu_g)\} \right], \quad (7)$$

where L is a function of the elements of w_g, $\mathbf{B}_g$, Γ^g, Λ_x^g, Λ_y^g, ν_x^g, ν_y^g, α^g, Φ^g, ψ^g, Θ_δ^g, Θ_ε^g, and τ_ξ^g for $g = 1, \ldots, G$. The problem is to maximize L with respect to the free parameters given the sample data $(\Delta_1, \ldots, \Delta_N)$ and a specified number of groups G, while taking into account the constraints imposed on $\mathbf{w}$ above, and $|\Sigma_g| > 0$ for all g.[8] Note that the maximum likelihood estimates $\hat{\Sigma}_g$ and $\hat{\mu}_g$ are functions of the postulated theoretical model: the measurement model *and* the structural model (see Equations (4) and (5)), as well as the mixing proportions $\hat{w}_g$.

We develop a maximum likelihood procedure to maximize Equation (7), using a modified *E-M* algorithm (see Dempster, Laird, and Rubin 1977). (Mathematical derivations and details regarding the algorithm are described in a working paper available from the first author.) The *E-M* method provides two important advantages. First, the method gives monotone increasing values of the log-likelihood function (see McLachlan and Basford 1988, p. 15 and Titterington, Smith, and Makov 1985, p. 85). Hence, we can show convergence to at least a locally optimum solution using a limiting sums argument, assuming that the estimated covariance matrices are nonsingular (Courant 1953, pp. 370–371). Second, based on our experience, the algorithm appears to be reasonably robust over the choice of initial starting values for the parameters. (This issue will be discussed further in §§5 and 6.) A disadvantage of the *E-M* algorithm is that the convergence rate can be slow (McLachlan and Basford 1988, p. 17).

Upon achieving convergence, the algorithm provides estimates of the model parameters and their asymptotic covariances (see McLachlan 1992, pp. 41–43). It is straightforward to use the parameter estimates to assign individual observations, i, to each of the G segments by using Bayes' rule:

$$\hat{P}_{ig} = \frac{\hat{w}_g f_g(\Delta_i | \hat{\Sigma}_g, \hat{\mu}_g)}{\Sigma_{k=1}^{G} \hat{w}_k f_k(\Delta_i | \hat{\Sigma}_k, \hat{\mu}_k)}, \quad (8)$$

where $\hat{P}_{ig}$ denotes the posterior probability of observation i belonging to segment g. These posterior probabilities represent a fuzzy classification (clustering) of the N observations into G segments conditional on the structural equation model postulated.

In practice, as in any other structural equation model estimation procedure, it is necessary to apply the *E-M* algorithm several times using a wide range of starting values. This step is important because it is necessary to search for all local maxima and choose the highest of these values. McLachlan and Basford (1988, pp. 17–18, and pp. 38–40) discuss the relevant statistical theory succinctly, including the case where the likelihood function is unbounded (e.g., the covariance matrices differ across groups).

Finally—and especially in consumer behavior studies—one can interpret the G groups as defining a discrete unobserved moderator of the relationships among the dependent and independent constructs in the structural equation model. Once the data have been partitioned into the G groups, the researcher can im-

[8] The condition $|\Sigma_g| > 0$ requires a minimum sample size of $(p + q)(p + q + 1)/2$ for each group where p and q, respectively, denote the numbers of indicators for the endogenous and exogenous constructs. This condition should be satisfied in most marketing applications. As in the standard multigroup structural equation model, empirical underidentification can occur if any group is small.

Copyright © 2001 All Rights Reserved

prove marketing policy by performing a posterior analysis to relate segment membership to observed background variables (e.g., demographics and psychographics).

Model Selection

Model selection depends on the researcher's *a priori* knowledge of the problem structure. Suppose the researcher knows the number of segments *a priori*. Then he or she can compare alternative nested models directly using the likelihood ratio statistic because the regularity conditions hold. In most practical cases, however, we expect the researcher to use the finite mixture structural equation model in applications where the number of segments G is not known *a priori*. It is therefore necessary to choose statistical criteria for determining the optimal number of segments. In contrast to the standard multigroup case, conventional tests based on the likelihood ratio (e.g., the chi-squared statistic) do not apply because the regularity conditions are violated. See Titterington, Smith, and Makov (1985, pp. 4–5) for a simple example. This problem is widely recognized in the literature and a variety of heuristics has been proposed (see McLachlan and Basford 1988, pp. 35–36). In keeping with the spirit of finite mixture analysis, we propose that the researcher choose the number of segments by examining several global measures of fit such as the consistent Akaike Information Criterion CAIC (Bozdogan 1987) and the Bayesian Information Criterion BIC (Schwarz 1978).

Assuming that a G-segment model is satisfactory, it is useful to assess the degree of separation among the various segments. We use an entropy-based measure E_G (Ramaswamy, DeSarbo, Reibstein, and Robinson 1992) based on the posterior probabilities. Specifically, E_G is defined by:

$$E_G = 1 - \left[\sum_i \sum_g - \hat{P}_{ig} \ln \hat{P}_{ig}\right] \Big/ (N \ln G), \qquad (9)$$

where $\hat{P}_{ig}$ is defined in Equation (8). This measure is bounded by 0 and 1. A value close to 0 indicates that the posterior probabilities are not well separated (i.e., it is difficult to classify observations accurately into distinct groups).

5. A Customer Satisfaction Application

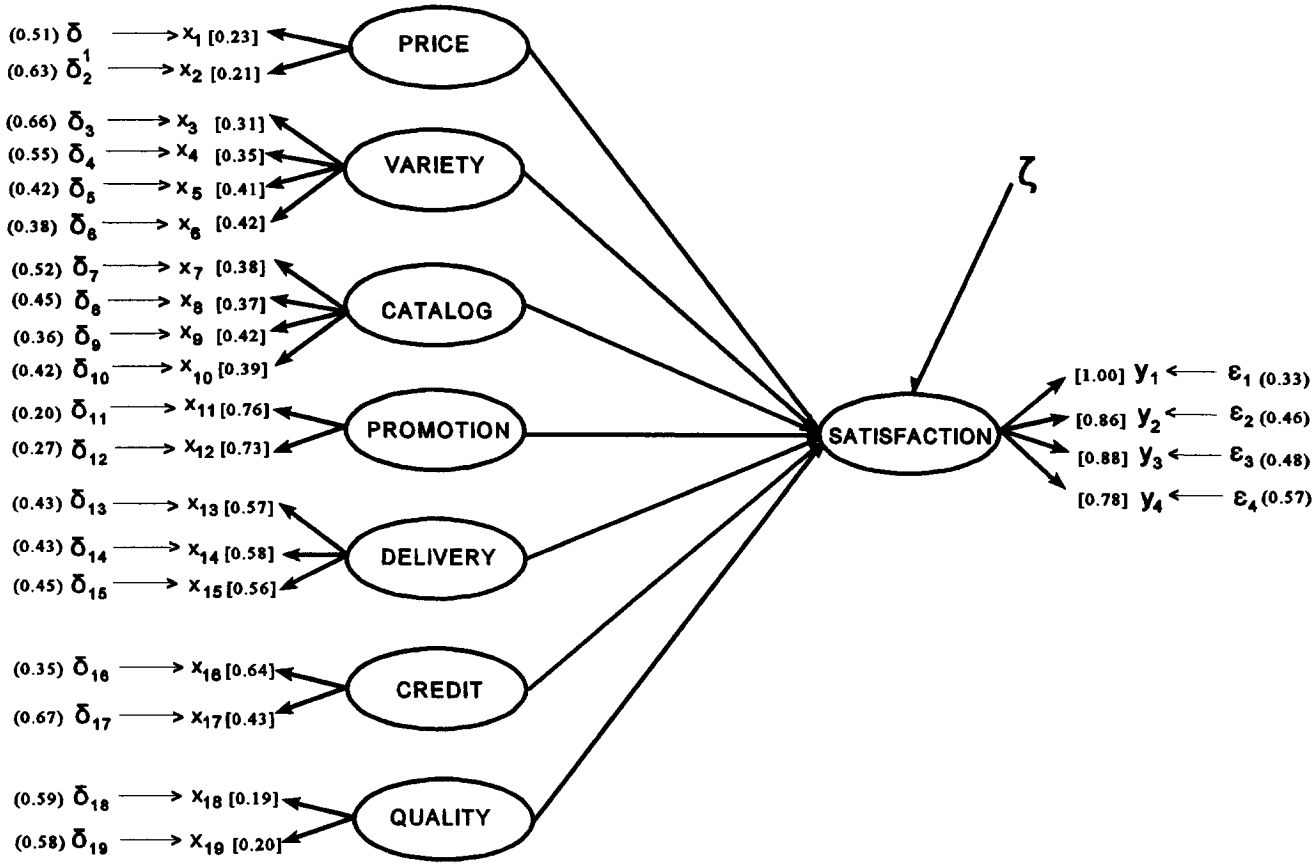
Several researchers have used structural equation models to analyze the antecedents of consumer satisfaction (e.g., Fornell and Westbrook 1984). We applied the finite mixture structural equation method to a satisfaction model using data from a customer satisfaction study conducted for a home shopping club in Europe. The data were collected by a European direct marketing firm using a telephone survey. The sample size was 1,564 customers. After discussing the substantive issues with the client and examining several exploratory analyses, we hypothesized that seven latent dimensions determine overall satisfaction (η). They are satisfaction with price, variety, catalog, promotion, delivery, credit, and quality. The observable indicators for price (ξ_1) are satisfaction with delivery charge (x_1) and with overall price (x_2). The observable indicators for variety (ξ_2) are satisfaction with overall product range (x_3), brands (x_4), fashion (x_5), and exclusivity (x_6). The indicators for catalog (ξ_3) are satisfaction with overall catalog (x_7), picture quality (x_8), information (x_9), and style (x_{10}). The indicators for promotion (ξ_4) are satisfaction with letters (x_{11}) and with gifts (x_{12}). The indicators for delivery (ξ_5) are satisfaction with availability (x_{13}), shipment (x_{14}), and speed (x_{15}). The indicators for credit (ξ_6) are satisfaction with the procedure (x_{16}) and with reminders (x_{17}). The indicators for quality (ξ_7) are satisfaction with general quality (x_{18}) and with product guarantees (x_{19}). The indicators for overall satisfaction (η) are general satisfaction (y_1), closeness to ideal outlet (y_2), conformance (y_3), and care (y_4). All variables were measured on ten-point scales where higher scores denote higher levels of the variables. See Figure 2 for a summary of the causal structure postulated.

Two research strategies are possible. First, one can estimate a single-group structural equation model using the aggregate data. This approach implicitly assumes that the data are homogeneous. Second, one can theorize that there are distinct segments each of which possibly follows a different model.[9] For example,

[9] Alternatively, one can argue that a discrete unobserved moderating factor accounts for the heterogeneity.

Copyright © 2001 All Rights Reserved

Figure 2 **Theoretical Model for Customer Satisfaction Study***



*Notes:

1. Numbers in parentheses denote the estimated variances of the disturbance terms.
2. Numbers in square brackets denote factor loading estimates.
3. All the variance and factor loading parameters are significant at $p < 0.01$ level

consumers could belong to the "highly satisfied," "highly unsatisfied," or "moderately satisfied" segments and so on (i.e., the factor means could differ across segments). The underlying facets of satisfaction could have differential effects (i.e., the structural parameters could differ across segments). The accuracy with which respondents answer questions could also depend on task involvement (i.e., the measurement parameters could vary by segment).

In order to test for unobserved heterogeneity, we analyzed the aggregate and finite mixture structural equation models. To reduce the chance of local optima, we estimated each model ten times using different random starting values.

The results for the aggregate data ($G = 1$ solution) suggest that except for price all the facets of satisfaction are important (see Table 2). Before accepting these results we tested for unobserved heterogeneity using the following general measurement and structural models:

$$\mathbf{x}_i \mid g = \boldsymbol{\nu}_x^g + \boldsymbol{\Lambda}_x^g \boldsymbol{\xi}^g + \boldsymbol{\delta}^g, \tag{10a}$$

$$\mathbf{y}_i \mid g = \boldsymbol{\nu}_y^g + \boldsymbol{\Lambda}_y^g \boldsymbol{\eta}^g + \boldsymbol{\varepsilon}^g, \tag{10b}$$

$$\mathbf{B}_g \boldsymbol{\eta}^g = \boldsymbol{\alpha}^g + \boldsymbol{\Gamma}^g \boldsymbol{\xi}^g + \boldsymbol{\zeta}^g. \tag{10c}$$

Copyright © 2001 All Rights Reserved

Table 1 Summary Statistics for Model Selection in Satisfaction Study

G	Measurement Model	M_G	$-\text{Ln } L$	CAIC	BIC	GFI	RMR	E_G
1	–	76[a]	43950.7	88536.3	88460.3	0.79	0.17	–
2	invariant	92	42958.0	86684.7	86592.6	0.89	0.06	0.78
2	free	114	42866.3	86685.1	86571.1	0.89	0.05	0.80
3	invariant	108	42793.2	86488.8*	86380.7*	0.89	0.05	0.70
3	free	152	42637.4	86544.7	86392.7	0.90	0.05	0.72
4	invariant	124	42754.9	86545.9	86421.8	0.89	0.05	0.62
4	free	190	42551.4	86690.3	86500.3	0.89	0.06	0.62

* Denotes minimum values for CAIC and BIC.
[a] M_G denotes the number of free parameters.

Equations (10a), (10b), and (10c) imply that the means and reliabilities of all latent variables differ across segments, the intercept terms in the structural equation model vary across groups, and the differential impacts of the dimensions of satisfaction on overall satisfaction vary by group.

To achieve model identification, we impose the following constraints: $E(\boldsymbol{\xi}^1) = \mathbf{0}$ (the null vector), $\boldsymbol{\alpha}^1 = \mathbf{0}$, $\text{Var}(\xi_i^g) = \phi_{ii}^g = 1$ for all i, g, and $\lambda_{y_1}^g = 1$ for all g. Hence $\boldsymbol{\tau}_\xi^g (g > 1)$ measures the differences in the factor means across groups for the exogenous latent variables, where Segment 1 is the reference group.

We estimated several finite mixture structural equation models, varying the number of segments and allowing for general types of heterogeneity (i.e., we allowed for different structural and measurement models across segments). Note that although we could have simply asserted that the measurement model is invariant across groups, we chose a more conservative strategy because of the lack of previous research using our method. Recall that we can also interpret "segments" as discrete unobserved moderators of the relationships among the constructs in the structural equation model.

The results (see Table 1) show that both the BIC and CAIC criteria point to the three-segment ($G = 3$) solution in which the measurement model is invariant across groups. This result of measurement model invariance suggests that the indicators are well chosen.

We now analyze in detail the results for the three-segment solution with an invariant measurement model. The results (i.e., factor means and structural parameters) for the overfitted three-segment solution with heterogeneous measurement models are very similar and are omitted.

The results of the finite mixture analysis suggest that the data are heterogeneous. All the reliabilities in the three-group solution have significant loadings, suggesting a good fit (see the factor loading estimates in Figure 2). The mean levels for the satisfaction constructs across groups are given in the first panel of Table 2. The results show that Segment 3 (17.3% of the sample) has the highest mean satisfaction levels across all constructs. Segment 2 (39.0% of the sample) has the lowest mean satisfaction levels for all constructs. Furthermore, the mean differences for all latent variables are highly significant across all groups. Hence, the manager might consider steps to improve the satisfaction scores for Segment 2. Before modifying the marketing mix, however, it is necessary to examine the structural model for each segment in detail.

Panel 2 of Table 2 gives the estimates of the structural parameters for the aggregate model and the three-segment solution. As discussed previously, the aggregate model ($G = 1$) suggests that all dimensions except price are significant determinants of customer satisfaction. The finite mixture results, however, show that there are both common and segment-specific effects across the three market segments. Only delivery, credit, quality, and promotion influence overall customer satisfaction in Segment 1. In contrast, catalog, promotion, delivery, credit, and quality determine overall customer satisfaction in Segment 2. Only delivery and credit significantly impact overall customer satisfaction in Segment 3.

Copyright © 2001 All Rights Reserved

Table 2 Parameter Estimates for the Aggregate and Three-Segment Solutions in the Satisfaction Study

			Structural Parameters			
	Factor Mean Scores			Finite Mixture		
Factor	Segment 2[a]	Segment 3	Aggregate Solution	Segment 1	Segment 2	Segment 3
η (Satisfaction)	−0.6	0.85	–	–	–	–
ξ_1 (Price)	−2.53**	2.77**	0.060	0.093	−0.006	0.011
	(.487)[b]	(.515)	(.037)	(.068)	(.063)	(.095)
ξ_2 (Variety)	−1.27**	1.84**	0.074*	0.031	0.051	0.061
	(.115)	(.155)	(.033)	(.032)	(.034)	(.059)
ξ_3 (Catalog)	−1.56**	1.87**	0.237**	0.058	0.275**	0.115
	(.118)	(.160)	(.036)	(.036)	(.032)	(.062)
ξ_4 (Promotion)	−0.76**	0.62**	0.077*	0.075*	0.079**	0.000
	(.084)	(.100)	(.032)	(.027)	(.026)	(.037)
ξ_5 (Delivery)	−0.86**	1.07**	0.207**	0.183**	0.188**	0.163**
	(.092)	(.137)	(.036)	(.030)	(.025)	(.049)
ξ_6 (Credit)	−0.83**	0.58**	0.170**	0.167**	0.119**	0.163**
	(.102)	(.171)	(.040)	(.035)	(.028)	(.055)
ξ_7 (Quality)	−2.967**	3.96**	0.234**	0.331**	0.213**	0.099
	(.635)	(.814)	(.049)	(.047)	(0.062)	(.117)
α	–	–	–	0	0.830	−0.171
					(.348)	(.615)
Mixing Proportions	–	–	–	0.437	0.390	0.173

[a] Segment 1 is the reference group (i.e., all factor means for segment 1 are set to zero).
[b] Standard errors are in parentheses.
** Denotes significant at the 0.01 level.
* Denotes significant at the 0.05 level.

These results show that aggregate analysis is misleading. They also show that, even if the confirmatory factor model for all the exogenous variables were homogeneous, the structural model would be misspecified if we performed an aggregate analysis using the pooled data.[10]

In practice, the next step would be to use the posterior probabilities from the finite mixture model to classify individuals into the three segments. These results could be used to improve marketing efficiency by relating group membership to such individual-specific variables as demographic and psychographic characteristics. Lacking such data, we could not perform this posterior analysis.

[10] It is straightforward to verify that, regardless of which dimension (ξ_i) of satisfaction we consider, the 99% confidence intervals for the means across groups do not overlap (see Table 2). Hence pooled data analysis is not appropriate.

Cross-Validation

We performed a double cross-validation analysis to check the stability of the results. This method requires one to randomly split the sample into halves, estimate the model separately for each subsample, and use the estimates from one subsample to predict (validate) the outcomes for the other. The correlation between the vector of parameter estimates from the first subsample and the corresponding vector from the full sample is COR = 0.985 and the mean absolute deviation for the two sets of coefficients is MAD = 0.034. The corresponding values for the second subsample are COR = 0.987 and MAD = 0.030. These double cross-validation results reinforce the validity of our conclusions (i.e., the estimates are stable).

Copyright © 2001 All Rights Reserved

Comparison to Sequential Data Analysis

We compared the finite mixture structural equation method to a sequential data analysis strategy: performing K-means cluster analysis of the observed variables followed by multigroup structural equation modelling for $G = 3$ and an invariant measurement model. We also performed a double cross-validation analysis for both methods. The finite mixture method consistently outperformed the sequential data analysis strategy regardless of the choice of model fit statistics or validation criteria. For example, CAIC = 87155.2 and BIC = 87047.1 for the K-means model and CAIC = 86488.8 and BIC = 86380.7 for the finite mixture model. Detailed results are available from the first author.

6. Testing the Finite Mixture Model: Simulation Evidence

We performed four simulation experiments to test the robustness of the finite mixture model. The first two experiments examine the performance of different estimation methods for correctly specified models; the other two experiments focus on the effect of model misspecification. The first experiment examines the robustness of different statistical criteria (e.g., BIC and CAIC) for model selection in the finite mixture model. The second experiment compares the performances of the finite mixture method and a sequential data analysis strategy (K-means clustering followed by multigroup structural equation modeling). The third experiment examines the effect of estimating a finite mixture structural equation model when the correct model has random coefficients. The fourth experiment examines the effect of distributional misspecification on the performance of the finite mixture structural equation model.

Most simulation studies use a fixed set of parameter values. Thus the results are likely to depend heavily on the particular set of parameters chosen. To avoid this difficulty, we used *separate* sets of parameter values for all replications and treatments. Appendix 2 provides the details of how we generated the data in the four simulation studies. For each treatment condition we performed five replications and estimated the model ten times to avoid local optima. In all four experiments, we used a moderately large fixed sample size of 1,500.

Model Performance Criteria

The model performance criteria of interest are: goodness-of-fit, group separation, recovery of group membership, recovery of the measurement model, recovery of the structural model, and recovery of the full model. We used several measures of each criterion. We used RMR and GFI as measures of goodness-of-fit and entropy (see Equation (9)) as a measure of group separation. We used two measures for the recovery of group membership. The first was a matching coefficient MATCH (a hit rate based on classifying an observation into the segment for which the posterior probability is the highest); the second measure, CORP, was the biserial correlation coefficient between the estimated posterior probabilities and the true group memberships. We used two measures for the recovery of the measurement (structural) model. The first measure is a normalized root mean square statistic, RMSM (RMSS). This measure is scaled such that a value of one means perfect parameter recovery.[11] The second measure, CORM (CORS), is the correlation coefficient between the true measurement (structural) parameters and their estimates. We measured the recovery of the full model (measurement and structural models) similarly using RMSMS and CORMS. We transformed the response variables where necessary to avoid heteroscedasticity in the ANOVA analysis. Specifically, we applied Fisher's Z transform to all correlations and used the log-odds transformation in cases where the measure varied between zero and one. Because of space constraints, we summarize the main results. Details are available from the first author.

First Simulation: Model Selection

Our primary purpose in this experiment was to examine the abilities of different model selection criteria (e.g., CAIC, BIC, GFI, RMR, and Entropy) to choose the correct number of clusters for exactly identified and overidentified models, assuming that the distributional form

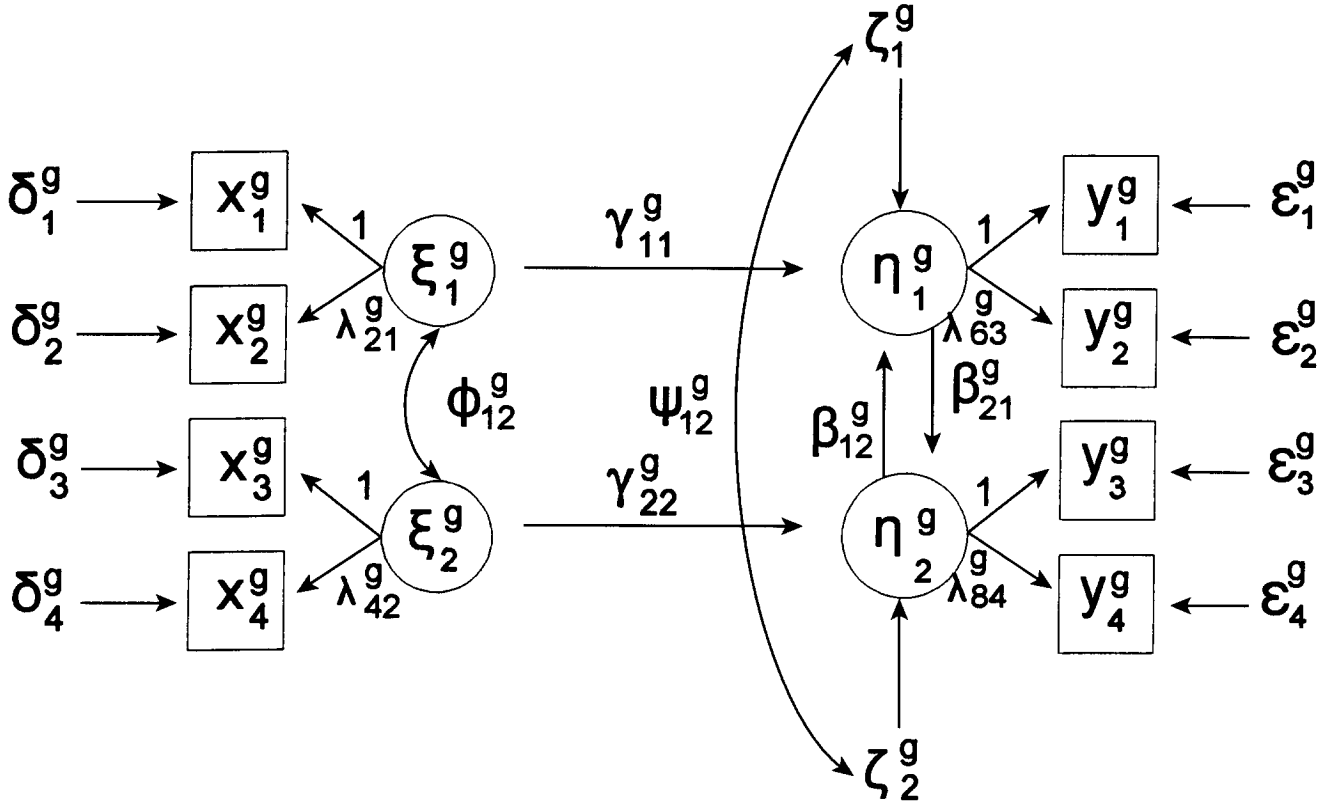
[11] Let λ_i^g denote any measurement parameter for group g. Then

$$\text{RMSM} = 1 - \left(\sum_g \sum_i (\lambda_i^g - \hat{\lambda}_i^g)^2 \Big/ \sum_g \sum_i (\lambda_i^g)^2 \right).$$

A similar definition applies for the root mean square statistic RMSS of the structural model.

Copyright © 2001 All Rights Reserved

Figure 3 **Theoretical Model for Simulated Data: Two-Group, Nonrecursive Model**



(i.e., multivariate normality) is correctly specified. We estimated a one-factor model using a (2 × 2) factorial design with the following treatments: number of clusters (2 and 4) and number of indicators (3 and 6). To assess the abilities of the different information criteria (e.g., BIC) in picking the correct number of groups, we estimated the finite mixture model by varying the number of groups from one through six.

The results show that only BIC and CAIC are effective for choosing the number of groups. For example, in the four-group model, the average GFI statistics across simulation runs are, respectively, 0.96, 0.95, 0.98, 0.99, and 0.99 for the one-, two-, three-, four-, and five-group cases. The corresponding estimates for RMR are 0.05, 0.06, 0.06, 0.028, and 0.026.

Overall, the BIC criterion was somewhat superior to CAIC and picked the correct number of groups in 92.5% of the cases. In contrast, the success rate for CAIC was 85%. Both BIC and CAIC performed extremely well for the overidentified six-indicator models. The success rate for BIC was 100% for both the two- and four-group models. The success rate for CAIC ranged from 90% in the four-group model to 100% in the two-group model. Both BIC and CAIC performed less well in the exactly identified three-indicator models. In particular, the success rate for each criterion fell when the number of clusters increased. The success rate of BIC (CAIC) fell from 90% (90%) in the two-group model to 80% (60%) in the four-group model. Thus both BIC and CAIC perform well for overidentified models. These results are encouraging because most structural equation models are overidentified for theoretical reasons. BIC, however, appears to be a superior model selection criterion for exactly identified models.

Second Simulation: Recovery of Parameters and Segment Membership

This experiment examines the nonrecursive model shown in Figure 3. The primary focus is on comparing

Copyright © 2001 All Rights Reserved

the abilities of the finite mixture model and a two-step approach (*K*-means clustering followed by multigroup structural equation analysis) to recover segment membership and, for each method, to determine the parameter biases for different model specifications.[12] (We used PROC FASTCLUS in SAS to perform the *K*-means clustering.) We also measured the extent of bias in the parameter estimates produced by an aggregate analysis which ignores heterogeneity.

We chose a (4 × 2) factorial design and used the following treatments: estimation method (aggregate AGG; known group membership MG; finite mixture FM; and *K*-means KM) and model specification (i.e., a model with invariant measurement loadings and free structural parameters across groups and a model with all measurement and structural parameters freely varying across groups). We included two covariates. The first covariate is the variance of the mixing proportions (VLAM) and is used to capture the effect of dispersion in group sizes. The second covariate is the average generalized Mahalanobis distance (DIST) between all pairs of population group centroids. We chose DIST because this measure explicitly allows for unequal covariance matrices across groups to capture the effect of group separation. In this experiment, the number of groups was fixed to two and was treated as known *a priori* in the estimation.

Table 3 shows the cell means for each performance criterion for different model specifications and method treatments. To facilitate comparison, consider the known group membership (MG) case as the benchmark. The finite mixture (FM) method performs extremely well in recovering group membership: the *K*-means (KM) results are less satisfactory. For example, in the treatment with free measurement and structural parameters, the CORP statistics for FM and KM are 0.945 and 0.840, respectively. The results for parameter bias show that, regardless of which set of parameters we consider (i.e., the measurement model, the structural model, or the complete model) or which performance criterion we use, the known group membership (MG) and finite mixture (FM) models significantly outperform the aggregate and *K*-means models (AGG and KM). For all comparisons and choices of performance criteria, however, the FM method performs as well as the benchmark model MG, i.e., the results are not significantly different.[13]

Note that for the aggregate model even though the standard goodness-of-fit statistics (e.g., GFI) are excellent, the recovery of both the measurement and structural parameters is extremely poor. Thus hypothesis testing using aggregate analysis is seriously misleading.

In order to determine the joint effects of the treatments, we performed an analysis of variance for each response variable (e.g., CORP) using the treatments and their interactions as independent variables. The last column of Table 3 shows the overall fit statistics for each response variable. Regardless of which performance criterion we choose, the main effects of the estimation method are highly significant ($p < 0.001$ in all cases except for RMR, for which $p < 0.014$). The main effects of the type of model (fixed or free measurement models across groups) are not significant. For the goodness-of-fit statistics GFI and RMR there is a significant interaction between estimation method and type of model ($p < 0.0001$ and $p < 0.047$, respectively). The group membership recovery statistics CORP and MATCH improve when the generalized Mahalanobis distance DIST is high ($p < 0.003$ and $p < 0.001$, respectively).

These results show that the finite mixture method (FM) performs well. To further examine the effect of group separation, we performed a median split on DIST. The *K*-means model (KM) is unsatisfactory when

[12] To our knowledge, this study is the first to examine the robustness of different methods (e.g., *K*-means clustering) where there is an explicit model structure and measurement error is present. Previous studies which examine measurement error do not allow for model structure; those studies which incorporate model structure do not allow for measurement error. See the review papers by Milligan and Cooper (1987) and, in particular, Bock (1996, pp. 404–405). In our simulations we used a sequential data-analysis strategy. We first performed *K*-means clustering to determine the appropriate number of groups and the corresponding group membership. We then applied multigroup structural equation modeling on the partitioned data. Alternatively, we could have estimated the multigroup structural equation model for different numbers of clusters and chosen that solution which provides the best fit in terms of the CAIC or BIC criteria.

[13] We performed pairwise comparisons using orthogonal contrasts in the analysis of variance.

Copyright © 2001 All Rights Reserved

Table 3 Summary Measures for the Accuracy of Different Estimation Methods: Nonrecursive Model

Measures		Measurement Model: Invariant MG	Invariant FM	Invariant KM	Invariant AGG	Free MG	Free FM	Free KM	Free AGG	R^2
Goodness-of-Fit	GFI	0.992	0.993	0.958	0.997	0.993	0.993	0.987	0.990	0.65**
	RMR	0.048	0.043	0.176	0.029	0.035	0.032	0.049	0.042	0.46*
Measurement Model Recovery	RMSM	0.055	0.056	0.262	0.732	0.091	0.118	0.197	0.655	0.82**
	CORM	0.998	0.998	0.960	0.771	0.994	0.990	0.974	0.785	0.88**
Structural Model Recovery	RMSS	0.123	0.119	0.614	0.916	0.087	0.098	0.249	0.317	0.63**
	CORS	0.956	0.957	0.703	0.685	0.971	0.964	0.752	0.636	0.68**
Full Model Recovery	RMSMS	0.082	0.081	0.401	0.842	0.091	0.115	0.217	0.584	0.89**
	CORMS	0.995	0.995	0.903	0.693	0.992	0.987	0.958	0.771	0.83**
Group Membership Recovery	CORP	1.00[a]	0.960	0.840	–	1.00[a]	0.945	0.845	–	0.82**
	MATCH	1.00[a]	0.950	0.841	–	1.00[a]	0.936	0.848	–	0.76**

[a] These values are trivially equal to unity because group membership is known.
* Denotes significant at 0.05 level.
** Denotes significant at 0.01 level.
MG Denotes the case where group membership is known.
FM Denotes the finite mixture model where group membership is unknown.
KM Denotes the *K*-means model where group membership is unknown.
AGG Denotes the aggregate model.

the groups are not well separated (i.e., DIST is low). If the groups are well separated (i.e., DIST is high), both FM and KM perform equally well in recovering group membership; however, FM is significantly superior to KM in recovering both the structural and the measurement parameters. These results are consistent with previous empirical studies which find that even in the absence of measurement error, mixture models significantly outperform fixed-classification models (e.g., *K*-means clustering) in parameter recovery. See Bock (1996, pp. 404–405) for a succinct review.

Third Simulation: Robustness to Misspecification of Form of Heterogeneity

This experiment focuses on model specification bias where the true but unknown form of heterogeneity is a random coefficient structural equation model. Following Sörbom (1981, p. 194) we assume fixed measurement parameters. The structural coefficients, however, vary randomly across the population according to a multivariate normal distribution with a diagonal covariance matrix. Our primary goals were to examine the robustness of the aggregate, finite mixture, and *K*-means estimators and to compare the performance of different model selection criteria (e.g., BIC and CAIC). Recall that we cannot use a random coefficient structural equation estimator as a benchmark because this estimator has not been developed in the statistical and psychometric literature.

We examined a nonrecursive random coefficients structural equation model (see Figure 3) using a (3×2) factorial design with the following treatments: estimation method (aggregate AGG; finite mixture FM; and *K*-means KM) and level of heterogeneity in the random coefficients[14] (low and high).

[14] We operationalized the levels of heterogeneity in the random coefficient structural equation model using the following procedure. Let $\bar{\gamma}_{11}$, $\bar{\gamma}_{22}$, $\bar{\beta}_{12}$, and $\bar{\beta}_{21}$ denote the population means of the structural parameters (see Figure 3). Consider the structural equation $\eta_1 = (\bar{\gamma}_{11} + \epsilon_1)\xi_1 + (\bar{\beta}_{12} + \epsilon_2)\eta_2 + \zeta_1$ where ϵ_1 and ϵ_2, respectively, denote the random deviations of γ_{11} and β_{12} from their population means. Then $\epsilon_1\xi_1 + \epsilon_2\eta_2$ is the effect of parameter heterogeneity and ζ_1 the effect of random error. Let a_1 and a_2, respectively, denote the variances of $(\epsilon_1\xi_1$

Copyright © 2001 All Rights Reserved

Table 4 shows the parameter biases of the different estimation methods for each heterogeneity level. We report the biases of the finite mixture method when different information criteria (i.e., CAIC, BIC) are used for selecting the number of groups.

The *K*-means model is highly sensitive to specification error introduced by random coefficients, especially when parameter heterogeneity is high. In all replications, regardless of the level of parameter heterogeneity, the *K*-means method picked models with four or more groups. In the high heterogeneity treatment, the *K*-means method picked six-group solutions in two of the five replications.

The finite mixture method is less sensitive to specification error resulting from random coefficients. Both BIC and CAIC perform approximately equally well. For the low heterogeneity treatment, BIC and CAIC point to the one-group solution in one replication and to the two-group solution in four replications. Not surprisingly, high heterogeneity has an adverse effect. For example, CAIC leads to three-group solutions in four replications and to a four-group solution in one replication.

In order to obtain additional insight, we examined for each cell the average value across replications of the largest mixing proportions, AVG MAX MIX (see Table 4). The results show that, regardless of the performance criterion (CAIC or BIC) or level of heterogeneity, the finite mixture method tends to lead to one large group. For example, in the low heterogeneity treatment, AVG MAX MIX is 0.89 for both CAIC and BIC. In contrast, the *K*-means method (KM) tends to lead to a large number of groups of approximately equal size. For example, in the low heterogeneity treatment, AVG MAX MIX for KM is only 0.30. These results (i.e., one large group and several small groups) suggest that the finite mixture model provides useful diagnostic information when the model is misspecified (i.e., the true model consists of one population with random coefficients).

Interestingly, the goodness-of-fit statistics for any given level of heterogeneity are similar for all estimation methods and do not provide diagnostic information (see Table 4). The aggregate model AGG, however, performs best and *K*-means is the least satisfactory. For example, in the low heterogeneity treatment, the GFI statistics are 0.9972 and 0.925, respectively, for the AGG and KM models.

The aggregate model captures the measurement parameters (i.e., the λ's, θ's, and ϕ's) and the population means of the structural parameters extremely well (see Table 4). The average normalized root mean squares for these parameters are 0.98 or better. The aggregate model, however, performs poorly for the structural errors. The average normalized root mean squares are 0.56 and 0.10, respectively, for the low and high heterogeneity treatments. These results are not surprising because fixed coefficient models force the error terms to pick up the unobserved heterogeneity in the parameters.[15]

All four methods capture the measurement parameters (i.e., the λ's and θ's) extremely well. For example, the normalized root mean square for λ in the low heterogeneity treatment ranges from 0.990 for *K*-means to 0.997 for the aggregate model. Compared to the aggregate and the finite mixture methods, however, the *K*-means method does very poorly in recovering the covariance matrix of the exogenous constructs Φ. For example, the normalized root mean square for Φ in the high heterogeneity treatment is 0.41 for *K*-means and 0.996 for the aggregate model.

The biases in the structural parameters depend crucially on the estimation method. For the structural parameters β and γ, the aggregate and finite mixture methods are highly similar and perform extremely well (i.e., the normalized measures are close to unity). In contrast, the *K*-means results are unsatisfactory: the normalized measures for β and γ for the low and high heterogeneity treatments, respectively, are 0.416 and 0.841.

$+ \epsilon_2\eta_2$) and ζ_1. Then $a_1/(a_1 + a_2) = 0.3$ defines the low heterogeneity and $a_1/(a_1 + a_2) = 0.7$ the high heterogeneity treatment. We followed the same procedure to define low and high heterogeneity for η_2.

[15] Consider the simple regression model $y = \alpha + \beta x + u$ where x and y are observables measured without error and α and β are random coefficients. Let $E(\alpha) = \bar{\alpha}$ and $E(\beta) = \bar{\beta}$. Then $y = \bar{\alpha} + \bar{\beta}x + u'$ where $u' = (\alpha - \bar{\alpha}) + (\beta - \bar{\beta})x + u$. The disturbance u' in the fixed coefficient model is heteroscedastic and confounds random error (u) and the effect of parameter heterogeneity (i.e., $(\alpha - \bar{\alpha}) + (\beta - \bar{\beta})x$). Maddala (1977, pp. 401–403) discusses estimation methods for this type of single-equation random coefficient regression model.

Copyright © 2001 All Rights Reserved

Table 4 Summary Results for the Random Coefficient Model: Nonrecursive Specification

		Finite Mixture		
	AGG	CAIC	BIC	*K*-Means
A. Low Heterogeneity				
Parameter Recovery:[1]				
Lambda	0.99714	0.99935	0.99935	0.99086
Theta	0.99394	0.99600	0.99600	0.97471
Phi	0.99650	0.99718	0.99718	0.45228
Beta/Gamma	0.98546	0.97923	0.97923	0.41573
Psi	0.55823	0.84892	0.84892	0.52746
Goodness-of-Fit:				
GFI	0.9972	0.9952	0.9952	0.92460
RMR	0.0122	0.0254	0.0254	0.08620
Entropy	–	0.5710	0.5710	1.00[2]
Avg. Max Mix	–	0.8885	0.8885	0.29780
B. High Heterogeneity				
Parameter Recovery:				
Lambda	0.99827	0.99956	0.99953	0.99805
Theta	0.99518	0.99694	0.99658	0.94356
Phi	0.99606	0.97482	0.97792	0.41651
Beta/Gamma	0.98369	0.97043	0.96714	0.84124
Psi	0.10000	0.42685	0.48152	0.26742
Goodness-of-Fit:				
GFI	0.99760	0.98880	0.98880	0.92380
RMR	0.01160	0.03720	0.04060	0.07700
Entropy	–	0.63080	0.57060	1.0000[2]
Avg. Max Min	–	0.70374	0.60000	0.29840

[1] Parameter recovery is measured using normalized values. A value of unity denotes perfect recovery.

[2] The value of Entropy is trivially 1.00 for *K*-Means.

The results for the structural error ψ show that the *K*-means model confounds parameter heterogeneity and errors-in-equations. The finite mixture method does much better in separating these effects. For example, in the low heterogeneity case and when CAIC is used for model selection, the normalized root mean square for ψ is 0.53 for the *K*-means model and 0.85 for the finite mixture model.

In order to determine the joint effects of the treatments, we performed an analysis of variance for each response variable using the treatments and their interactions as independent variables. In contrast to the previous experiments, we did not include covariates because heterogeneity is a treatment and the true model includes only one population.

The dependent variables are the biases in the following sets of parameters: λ, θ, ϕ, β and γ, and ψ (see Figure 3). For all sets of parameters, the main effects of estimation method are highly significant ($p < 0.002$). For the structural parameters β, γ, and ψ the main effects of heterogeneity are significant ($p < 0.01$). In addition, for β and γ there is a significant interaction effect between estimation method and level of heterogeneity ($p < 0.02$). These results for specification error show that all methods capture the measurement model well. The structural model, however, is most accurately captured by the finite mixture method; the *K*-means method performs poorly. In particular, only the finite mixture model provides useful diagnostic information when the model is misspecified.

Fourth Simulation: Robustness to Distributional Misspecification

This experiment focuses on distributional misspecification. Specifically, we examine a nonrecursive model (see Figure 3) for which the number of groups is known *a priori*. Our primary focus is on the ability of the finite mixture method to recover the true parameters and segment membership as we vary the number of segments and violate the distributional assumption of multivariate normality. We also examined in detail which factors affect the incidence of local optima.

We chose a ($2 \times 2 \times 2$) factorial design and used the following treatments: number of clusters (2 and 4), skewness (0 and 1), and kurtosis (0 and 2.75). Because the data generation process does not allow us to simulate a distribution with skewness 1 and kurtosis zero (see Fleishman 1978, Figure 1, p. 527), we chose a skewness of 0.75 when the kurtosis is zero. This design and choice of factor levels are consistent with those used in previous simulations (e.g., Muthén and Kaplan 1985).

Table 5 summarizes the performance of the finite mixture model for different specifications of skewness, kurtosis, and the number of clusters. As Table 5 shows, the goodness-of-fit statistics are excellent. GFI is very stable and ranges from 0.988 to 0.994. The normalized RMR is highly satisfactory but less stable and ranges from 0.015 to 0.026. For the two-group model, the accuracy in predicting group membership is very high. For example,

Copyright © 2001 All Rights Reserved

MATCH (i.e., the hit rate) ranges from 0.969 to 0.983. Accuracy in predicting group membership in the four-group model is somewhat lower. MATCH ranges from 0.899 to 0.941. The entropy measure shows that the groups are well separated: the values range from 0.870 to 0.951. Entropy, however, tends to fall as the number of clusters increases.

The finite mixture algorithm estimated the measurement model accurately for both the two-group and four-group models. The normalized RMSM ranges from 0.993 to 0.997. CORM—the correlation between the predicted and actual measurement parameters—is almost perfect and ranges from 0.977 to 0.998.

The normalized RMSS varies from 0.923 to 0.951 in the two-cluster model and is highly satisfactory. The structural parameters for the four-cluster model are less accurately measured, especially when skewness and kurtosis are high (i.e., 1 and 2.75, respectively). In this case, RMSS drops sharply to 0.705. CORS—the correlation between the predicted and actual structural parameters—shows the same pattern as RMSS. The deterioration of fit, however, is less extreme (e.g., CORS = 0.883 when skewness = 1 and kurtosis = 2.75). Similar results are obtained when all measurement and structural parameters are analyzed jointly.

The analysis of variance results show that, regardless of the dependent variable chosen, model performance depends on two factors only: the distances between the groups, DIST, ($p < 0.001$) and the number of groups ($p < 0.01$). All the other main effects (i.e., skewness and kurtosis) and all the two- and three-way interactions are insignificant. The finite mixture model performs best when the number of groups is small and the groups are well separated (see Table 5). This finding is not surprising because the sample size was fixed and the number of parameters increases almost linearly with the number of groups.

The finite mixture model—like the standard structural equation model—is subject to local optima problems. In our simulation, we encountered local optima in 38 out of 200 runs (see Table 5). The binary logit results show that the incidence of local optima depends significantly on the number of groups and the average distance between the groups. Specifically, the proportion of local optima increases from 9% in the two-group case to 29% in the four group-case.

Summary of Simulation Results

In summary, the simulation results show that when the data are heterogeneous such goodness-of-fit statistics as GFI, RMR, and ENTR are not useful in choosing the correct number of groups.[16] Aggregate analysis is seriously misleading and does very poorly in recovering both the measurement and structural parameters. For the finite mixture method both BIC and CAIC performed well as model selection criteria for overidentified models. BIC, however, appears to be superior for exactly identified models. The finite mixture methodology is reasonably robust except when skewness and kurtosis are severe. We did not encounter any problems of empirical underidentification or convergence for the sample size examined ($N = 1500$). The K-means method is not satisfactory, especially when the groups overlap substantially. If the true model is a random coefficients structural equation model, both the finite mixture (FM) and K-means (KM) models are misspecified. FM, however, is much less sensitive to this specification error than KM. Of all methods examined including the aggregate model, the FM method performed best in capturing heterogeneity in the structural model, even though the model is misspecified. Finally, only the finite mixture model provides useful diagnostic information when heterogeneity occurs because of random coefficients.

7. Summary and Conclusions

This paper develops a finite mixture structural equation methodology to detect and treat unobserved heterogeneity. The method is appropriate for response-based market segmentation in the presence of measurement error and behavioral studies which postulate that unobserved moderating factors account for consumer heterogeneity.

Our method is general and subsumes a variety of statistical methods as special cases. These include finite mixture models in the frameworks of confirmatory factor analysis, simultaneous equation models with no

[16] This finding is not surprising because these fit measures are based on the elements of the aggregate covariance matrix which are not sufficient statistics for the model parameters (see Everitt and Hand 1981, p. 12).

Copyright © 2001 All Rights Reserved

Table 5 Summary Measures for the Accuracy of the Finite Mixture Algorithm: Nonrecursive Model

Measures		Number of Clusters: 2, Skew = 0, Kurtosis = 0	2, Skew = 0.75, Kurtosis = 0	2, Skew = 0, Kurtosis = 2.75	2, Skew = 1.0, Kurtosis = 2.75	4, Skew = 0, Kurtosis = 0	4, Skew = 0.75, Kurtosis = 0	4, Skew = 0, Kurtosis = 2.75	4, Skew = 1.0, Kurtosis = 2.75	Overall	R^2
Goodness-of-Fit	GFI	0.994	0.993	0.994	0.994	0.988	0.990	0.988	0.988	0.991	0.711**
	RMR	0.019	0.018	0.015	0.018	0.025	0.021	0.026	0.023	0.021	0.356
Group Separation	Entropy	0.947	0.906	0.951	0.954	0.918	0.870	0.897	0.903	0.918	0.672**
Group Membership	CORP	0.985	0.972	0.988	0.986	0.950	0.933	0.923	0.931	0.958	0.768**
Recovery	MATCH	0.983	0.969	0.986	0.982	0.941	0.914	0.899	0.928	0.950	0.778**
Measurement	RMSM	0.996	0.996	0.996	0.997	0.994	0.995	0.993	0.994	0.995	0.328
Model Recovery	CORM	0.998	0.998	0.998	0.998	0.997	0.997	0.997	0.997	0.997	0.332
Structural Model	RMSS	0.947	0.951	0.939	0.923	0.873	0.847	0.853	0.705	0.880	0.226
Recovery	CORS	0.976	0.981	0.972	0.966	0.945	0.944	0.936	0.883	0.950	0.344
Full Model Recovery	RMSMS	0.994	0.993	0.993	0.994	0.991	0.991	0.990	0.987	0.992	0.253
	CORMS	0.997	0.996	0.996	0.997	0.996	0.995	0.995	0.993	0.996	0.149
Local Optima*	LOC	0.120	0.040	0.120	0.080	0.360	0.280	0.240	0.280	0.190	0.190**

Notes:

* Logit analysis was performed because the dependent variable is binary. R^2 should be read as U^2.

** denotes significant at 0.01 level.

Copyright © 2001 All Rights Reserved

measurement error in variables, and second-order factor analysis. Simulation results suggest that the algorithm is reasonably robust even when a moderate sample size is used. In particular, CAIC and BIC statistics perform well in model selection provided the structural equation model is correctly chosen, the data for each segment follow a conditional multivariate normal distribution, and the models are overidentified (as is usually the case).

We applied the finite mixture method to a direct marketing study of customer satisfaction and fitted a large model with eight latent variables and 23 observable indicators. The results show that aggregate analysis can be seriously misleading when there are significant differences in model structure across segments. The finite mixture model appears to be robust (i.e., the results hold under cross-validation); sequential data analysis (i.e., *K*-means clustering followed by multigroup structural equation modeling) is less satisfactory.

The results show that, from a practical viewpoint, researchers should be alert to the presence of unobserved heterogeneity; in particular, standard goodness-of-fit measures in structural equation models do not provide diagnostic information. If the data belong to a finite number of groups and the groups are well separated, *K*-means clustering performs reasonably well in classifying observations; parameter recovery, however, is poor. If the groups are not well separated, *K*-means clustering performs poorly both in classification and in parameter recovery. In all cases (including misspecified models in which the true model has random coefficients), the finite mixture model performs well and provides useful diagnostic information. Thus, given that the researcher cannot determine the degree of separation among groups *a priori* and is primarily interested in accurately estimating the model structure, the prudent strategy is to use the finite mixture method.

Researchers should keep in mind the following caveats. The researcher should use only the finite mixture methodology when substantive theory supports the structural equation model formulation and *a priori* segmentation is infeasible. In particular, theory should be available suggesting that the data are heterogeneous and belong to a finite number of unobserved groups. We expect these conditions to hold in many marketing applications (e.g., clustering) where measurement error and heterogeneity coexist. The finite mixture method, like the standard multigroup structural equation model with known group membership, requires large samples. In particular, the method cannot be used to detect outliers or niche segments (i.e., segments for which the mixing proportions are close to zero). One practice that should be avoided is that of fitting a finite mixture model which is not well grounded in substantive theory and simply adding groups until a reasonable fit is found (Everitt and Hand 1981, p. 127). In particular, as in other statistical models, a good fit for the mixture model does not "prove" that the hypothesized causal structure is correct (Boulding and Staelin 1995).

Future research should focus on large-scale simulation studies to test the finite mixture method using a wide range of models and statistical distributions. From a substantive viewpoint it is necessary to apply our finite mixture method to a broad range of marketing applications where heterogeneity is typical (e.g., salesforce compensation plans and power relationships in channels of distribution). Theoretical research should extend the finite mixture model by allowing the mixing proportions to depend on prior information and/or subject-specific variables. Finally, in order to provide a fuller treatment of heterogeneity, we need to develop a general random coefficient structural equation model. Such a model is presently unavailable in the statistical and psychometric literatures.[17]

[17] The authors wish to thank Phipps Arabie, Mohini Jagpal, Alex Kogan, and Shaw Hwa Lo for their helpful comments and encouragement and Claes Fornell for his permission to use the satisfaction data. The authors also thank the Editor, Area Editor, and two anonymous reviewers for their constructive comments. Kamel Jedidi and Harsharanjeet Singh Jagpal, respectively, acknowledge the financial support of the Columbia Business School and the Faculty of Management, Rutgers University.

Appendix 1. The Identification of Finite Mixtures of Structural Equation Models

PROPOSITION. Suppose a given multigroup structural equation model for known groups is identified. Then the finite mixture of structural equation models is identified provided the data for all groups follow multivariate normal distributions.

PROOF. Consider any *G*-group finite mixture of structural equation models. Let $\Delta = (y, x)'$ denote the joint vector of observables where

Copyright © 2001 All Rights Reserved

y and x, respectively, denote the subvectors of observables for the endogenous and exogenous constructs, respectively. For group g let φ^g denote the parameter vector of the structural equation model, $\mu^g(\varphi^g)$ the implied mean vector of observables, and $\Sigma^g(\varphi^g)$ the implied covariance matrix of observables. Let Ω^g denote the vectors of all the unduplicated elements in μ^g and Σ^g, $\Omega = (\Omega^1, \ldots, \Omega^G)$ and $\varphi = (\varphi^1, \ldots, \varphi^G)$. By assumption, the structural equation model for known groups is identified. That is, the mapping $\varphi \to \Omega$ is one-to-one.

Let $F(\Delta, \varphi^g)$ denote the distribution function for the observables in group g and H the distribution function of the finite mixture of observables. Then the finite mixture structural equation model is identified if $H = \sum_{g=1}^{G} w^g F(\Delta, \varphi^g)$ implies that G, w^g, and φ^g are unique (see Titterington, Smith, and Makov 1985, p. 36).

Our proof proceeds by contradiction using the result that finite mixtures of multivariate normal distributions are identified (see Titterington, Smith, and Makov 1985, p. 162; Basford and McLachlan 1988, p. 97). Let $\varphi_1 = (\varphi_1^1, \ldots, \varphi_1^{G_1})$ and $\varphi_2 = (\varphi_2^1, \ldots, \varphi_2^{G_2})$ where G_1 and G_2 are not necessarily equal. Suppose the proposition is false. Then there must be some $\varphi_1 \neq \varphi_2$ such that

$$H = \sum_{g=1}^{G_1} w_1^g F(\Delta, \varphi_1^g) = \sum_{g=1}^{G_2} w_2^g F(\Delta, \varphi_2^g)$$

(see definition 3.1.1 in Titterington et al. 1985, p. 36).

Case 1: $G_1 = G_2$.

By assumption the mapping $\varphi \to \Omega$ is one-to-one. Hence $\varphi_1 \neq \varphi_2$ implies that $\Omega_1 \neq \Omega_2$ where Ω_1 and Ω_2 have the same dimensionality. Thus two finite mixtures of multivariate normal distributions with the same number of groups yield the same distribution function H. However, this contradicts the result that a finite mixture of multivariate normal distributions is identified. Hence $\varphi_1 = \varphi_2$.

Case 2: $G_1 \neq G_2$.

In this case, $\varphi_1 \neq \varphi_2$ implies that $\Omega_1 \neq \Omega_2$ where Ω_1 and Ω_2 have different dimensionalities. Thus H can be represented as a G_1-group or as a G_2-group finite mixture of multivariate normal distributions. This implication also contradicts the result that finite mixtures of multivariate normal distributions are identified. Hence $\varphi_1 = \varphi_2$, concluding the proof.

Appendix 2. Data Generation Procedure for Simulation Studies

We used *separate* sets of parameter values for all treatments and replications and a fixed sample size of 1,500.

In order to generate the mixing proportions, we chose random numbers a_g ($g = 1, \ldots, G$) from a uniform distribution with a range of 0 to 1. The mixing proportion for group g was defined as $w_g = a_g / \Sigma a_g$. In order to avoid undersized samples, we rejected any set of mixing proportions which led to any group with fewer than 200 observations. Thus, no group comprised less than 13.33% (i.e., 200/1500) of the population. Note that we tested the finite mixture model under stringent conditions; a sample size of 200 is small compared to the number of free parameters per group. Muthén and Kaplan (1985), for example, used a fixed set of parameters and a sample size of 1,000 in their simulation to test a single-group confirmatory one-factor model with only eight parameters.

In all experiments, we set one lambda per construct to unity for identification. For each remaining observable we generated the lambdas randomly from a uniform distribution with a range from 0.5 to 1.5. This procedure ensured that the lambdas varied freely across constructs and clusters. For each cluster g we chose the mean of each exogenous construct (ξ) randomly from a uniform distribution with range $-g$ to g. We chose these ranges to provide some degree of separation among the clusters. We set the variances of all exogenous constructs to unity for all clusters. Note that although we fixed these theoretical variances at unity for data generation, each variance is a free parameter in the model. For each indicator (including the reference indicators) we chose the reliability randomly from a uniform distribution with a range of 0.5 to 0.9. This range is consistent with typical findings in behavioral research. Having chosen a particular reliability, we determined the implied variance of the corresponding measurement error term.

For each group, we chose the correlation between the exogenous constructs randomly from a uniform distribution with a range of -0.8 to 0.8. For the structural model in Figure 3, we generated each beta and gamma randomly from a uniform distribution with a range from -1 to $+1$. For each structural equation in each group we chose R^2 randomly from a uniform distribution with a range from 0.3 to 0.8. This procedure determined the (random) variances of the structural error terms. For generality, we allowed the structural errors to be correlated (see Figure 3). We chose each such correlation coefficient randomly from a uniform distribution with a range from -0.8 to 0.8. We used these correlations to determine the covariances between the structural errors. For treatments involving skewness and kurtosis, we generated the data using the methods described in Fleishman (1978) and Vale and Maurelli (1984).

In all experiments, once a set of model parameters had been randomly generated for each group, we computed the implied group mean vectors and the implied covariance matrices and used these values to generate the sample data randomly.

References

Aitkin, M., D. Anderson, and J. Hinde (1981), "Statistical Modelling of Data on Teaching Styles (with Discussion)," *Journal of the Royal Statistical Society*, A 144, 419–461.

Bagozzi, Richard P. (1982), "A Field Investigation of Causal Relations Among Cognitions, Affect, Intentions, and Behavior," *Journal of Marketing Research*, 19, 562–584.

Baker, F. B. (1974), "Stability of Two Hierarchical Grouping Techniques Case 1: Sensitivity to Data Errors," *Journal of the American Statistical Association*, 69, 440–445.

Bock, H. H. (1996), "Probability Models and Hypothesis Testing in Partitioning Cluster Analysis," in Arabie, P., Hubert, L. J., and DeSoete G. (Eds.), *Clustering and Classification*, World Scientific Publishing Company, 377–453.

Bozdogan, H. (1987), "Model Selection and Akaike's Information Criterion (AIC): The General Theory and Its Analytical Extension," *Psychometrika*, 52, 345–370.

Copyright © 2001 All Rights Reserved

Boulding, W. and R. Staelin (1995), "Identifying Generalizable Effects of Strategic Actions on Firm Performance: The Case of Demand-side Returns to R & D Spending," *Marketing Science*, 14, 3, 222–236.

Chang, W. C. (1983), "On Using Principal Components Before Separating a Mixture of Two Multivariate Normal Distributions," *Applied Statistics*, 32, 267–275.

Chintagunta, Pradeep K., Dipak C. Jain, and Naufel Vilcassim (1991), "Investigating Heterogeneity in Brand Preferences in Logit Models for Panel Data," *Journal of Marketing Research*, November, 417–428.

Courant, R. (1953), *Differential and Integral Calculus*, Volume 1, 2nd ed. New York: Wiley.

Day, Ralph L. (1977), "Extending the Concept of Consumer Satisfaction," *in Advances in Consumer Research*, Vol. 4, W. D. Perreault (Ed.), Atlanta: Association for Consumer Research, 149–154.

Dempster, A. P., M. N. Laird, and D. B. Rubin (1977), "Maximum Likelihood Estimation from Incomplete Data via the *E-M* Algorithm," *Journal of the Royal Statistical Society, Series B*, 39, 1–38.

Everitt, B. S. and D. J. Hand (1981), *Finite Mixture Distributions*, New York: Chapman and Hall.

Fleishman, Allen A. (1978), "A Method for Simulating Non-normal Distributions," *Psychometrika*, 43, 521–532.

Fornell, Claes and Robert A. Westbrook (1984), "The Vicious Circle of Consumer Complaints," *Journal of Marketing*, 48, 68–78.

Jöreskog, Karl G. (1971), "Simultaneous Factor Analysis in Several Populations," *Psychometrika*, 36, 409–426.

Kamakura, Wagner A. and Gary J. Russell (1989), "A Probabilistic Choice Model for Market Segmentation and Elasticity Structure," *Journal of Marketing Research*, 26 (November), 379–390.

Kohli, Ajay K. (1989), "Effects of Supervisory Behavior: The Role of Individual Differences Among Salespeople," *Journal of Marketing*, 53, 40–50.

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

McLachlan, G. J. (1992), *Discriminant Analysis and Statistical Pattern Recognition*, New York: Wiley.

Milligan, G. W. and M. C. Cooper (1987), "Methodology Review: Clustering Methods," *Applied Psychological Measurement*, Vol. 11, No. 4, 329–354.

Moore, William L. (1980), "Levels of Aggregation in Conjoint Analysis: An Empirical Comparison," *Journal of Marketing Research*, 17, 516–523.

Muthén, Bengt O. (1989), "Latent Variable Modeling in Heterogeneous Populations," *Psychometrika*, 54, 557–585.

—— and David Kaplan (1985), "A Comparison of Some Methodologies for the Factor Analysis on Non-Normal Likert Variables," *British Journal of Mathematical and Statistical Psychology*, 38, 171–189.

Ramaswamy, Venkatram, Wayne S. DeSarbo, David J Reibstein, and William T. Robinson (1993), "An Empirical Pooling Approach for Estimating Marketing Mix Elasticities with PIMS Data," *Marketing Science*, 12, 1 (Winter), 103–124.

Schwarz, G. (1978), "Estimating the Dimension of a Model," *Annals of Statistics*, 6, 461–464.

Sörbom, Dag (1974), "A General Method for Studying Differences in Factor Means and Factor Structure between Groups," *British Journal of Mathematical and Statistical Psychology*, 27, 229–239.

Sörbom, Dag (1981), "Structural Equation Models with Structured Means," in Jöreskog, K. G. and H. Wold (Eds.), *Systems under Indirect Observation: Causality, Structure, and Prediction*, Amsterdam: North Holland.

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

Vale, C. D. and V. A. Maurelli (1983), "Simulating Multivariate Nonnormal Distributions," *Psychometrika*, 48, 465–471.

This paper was received January 18, 1996, *and has been with the authors* $2\frac{1}{2}$ *months for* 2 *revisions; processed by J. Wesley Hutchinson.*

Copyright © 2001 All Rights Reserved

Copyright 1997, by INFORMS, all rights reserved. Copyright of Marketing Science is the property of INFORMS: Institute for Operations Research and its content may not be copied or emailed to multiple sites or posted to a listserv without the copyright holder's express written permission. However, users may print, download, or email articles for individual use.