Journal of Classification 14:23-50 (1997)

STEMM: A General Finite Mixture Structural Equation Model

Kamel Jedidi

Columbia University

Harsharanjeet S. Jagpal

Rutgers University

Wayne S. DeSarbo

Pennsylvania State University

We wrote the first draft of this paper in May 1993. Since then we have revised the paper based on the insightful comments of Phipps Arabie, Michael W. Browne, Alex Kogan, and Michel Tuan Pham. We thank the editor and the anonymous referees for their constructive comments, and Richard Bagozzi for his permission in using the data from his panel study. The authors alone are responsible for all errors.

First author's address: Kamel Jedidi, Graduate School of Business, 518 Uris Hall, Columbia University, New York, NY 10027 USA.

Abstract: This paper provides a general STructural Equation finite Mixture Model and algorithm (STEMM). Substantively, the model allows the researcher to simultaneously treat heterogeneity and form groups in the context of a postulated causal (i.e., simultaneous equation regression) structure in which all the observables are measured with error. Methodologically, the model is more general than such statistical methods as cluster analysis, confirmatory multigroup factor analysis, and multigroup structural equation models. In particular the general finite mixture model includes, as special cases, finite mixtures of simultaneous equations with feedback, confirmatory factor analysis, and confirmatory second-order factor models. We describe the statistical theory, present simulation evidence on the performance of the EM estimation algorithm, and apply the model to a psychological study on the role of emotion in goal-directed behavior. Finally we discuss several avenues for future research.

Keywords: Structural equations; Finite mixtures; Maximum likelihood; Emotions; Weight loss

1. Introduction

Structural equation modeling is a powerful method for simultaneously estimating both the posited model structure linking unobserved constructs and the measurement errors in the observables (see Jöreskog 1971, 1973). If heterogeneity is present and group membership is known a priori, the researcher can use standard multigroup methods (Jöreskog 1971; Sörbom 1974). In practice, the researcher often does not have sufficient information to form groups a priori. Consequently as Muthén (1989, p.558) notes, "Data are frequently analyzed as if they were obtained from single populations although it is often unlikely that all individuals in our sample have the same set of parameter values." This homogeneity assumption is questionable at best. For example, in survey research the validities and reliabilities of items can vary across subgroups defined by race, gender, region, and issue salience (Muthén 1989, p.558). Similarly, in consumer psychology different consumer groups are likely to perceive brands differently depending on their previous purchase behaviors and media habits. Furthermore, while data on demographics, psychographics and other background variables are indeed typically collected by researchers, such information may be insufficient to allow researchers to form groups a priori (see Moore 1980).

This paper proposes a new STructural Equation finite Mixture Model (STEMM) and algorithm for handling heterogeneity in structural equation models which simultaneously forms groups and estimates the model structure and measurement errors for each group. The method is very general and subsumes finite mixture models for confirmatory factor analysis, second-order factor analysis, simultaneous equation models with feedback, and multiple regression as special cases. Furthermore, the method allows the researcher to use the mixture model in either a fully confirmatory or a partially

confirmatory mode. The former mode is appropriate when the researcher has a high degree of knowledge about the phenomena being analyzed (e.g., the researcher can use extant theory to specify the number of groups a priori or impose parametric restrictions based on previous empirical evidence). The partially confirmatory mode is appropriate when the researcher knows only the model structure but not the number of groups.

Section 2 discusses extant methods for treating heterogeneity, and Section 3 provides the technical details of our mixture model. Section 4 presents simulation results for three different model structures, and Section 5 applies our method to a study on goal-directed emotions. Finally, Section 6 discusses areas for future research.

2. Heterogeneity in Structural Equation Models

Previous research suggests two approaches for handling heterogeneity in structural equation models. One approach which assumes sufficiently large samples is to form groups a priori and then use the multigroup structural equation model (Sörbom 1974). A second approach which does not require such large samples or a priori grouping was recently proposed by Muthén (1989). This method allows the researcher to handle certain types of heterogeneity by using covariates and applying the MIMIC methodology to the pooled data. We discuss these approaches in turn.

Several methods are available for forming groups a priori. The researcher can apply some form of cluster analysis to the observed variables. 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.) In addition, different clustering procedures tend to yield different results, and little guidance is available on choosing the best procedure. Alternatively, 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 and then form clusters using the reduced dimensions. This approach is flawed because the first step assumes homogeneity (i.e., one population) whereas the second asserts heterogeneity (i.e., multiple populations). Empirical studies also show that these data-reduction strategies are not robust (see Chang 1983). See also McLachlan (1992, pp.196-203) for a succinct discussion of principal components and other data-reduction methods in the context of clustering.

The previous difficulties are compounded when the data include measurment error. Extensive simulation evidence shows that current clustering methods are highly unsatisfactory under these conditions. Baker (1974) found that classification accuracy was seriously affected when the data were

perturbed: the mean Goodman-Kruskal gamma statistics were as low as 0.365 for complete-link and 0.197 for single-link clustering for the high-error condition in his simulation experiments. Milligan (1980) compared 11 hierarchical and four non-hierarchical methods (e.g., K-means clustering) using cluster structures generated from truncated multivariate normal mixtures. He found that, despite the fact that the clusters were constructed to be non-overlapping, the introduction of measurement error led to a significant reduction in classification accuracy for all the algorithms tested. For details on the efficiency of clustering in the presence of measurement error, see the survey papers by Milligan (1981) and Milligan and Cooper (1987).

Finally, current clustering procedures do not postulate any causal (i.e., simultaneous equation regression) structure among the variables. This weakness is recognized in the literature. For example, McLachlan and Basford (1988, p. vi) note that their mixture method of clustering is meant for situations "where there is no a priori knowledge of any formal group structure in the underlying population, but where one wishes to cluster the data into a number of groups." Furthermore, McLachlan and Basford (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]. Hence we need a model-based clustering method which simultaneously allows both for causal relationships (i.e., a simultaneous equation regression structure) among the variables and for measurement error.

Muthén's (1989) alternative approach estimates one structural equation model using the pooled data, after including covariates to account for heterogeneity. His MIMIC method provides an important advantage: the researcher does not need to split the sample. However, this method requires the researcher to have sufficient a priori information or theory to specify covariates. Such information is not always available. Furthermore, as Muthén (1989, p. 564) notes, "In contrast to multiple-group analysis, the MIMIC approach is restricted to modeling under the assumption of a group-invariant covariance matrix for the observed response variables, conditional on grouping variables represented by the x's [covariates]. But with insufficient sample sizes for multiple-group analysis, this may be the best alternative." We agree that if the sample size is modest the researcher may have no alternative but to use the MIMIC approach. However, in many studies the sample size is reasonably large. Furthermore, for many real-life problems, the assumption of group-invariant covariance matrices is stringent. Hence, we require a large-sample method which allows for general types of heterogeneity and a causal structure recognizing both measurement and structural errors.

3. The Finite Mixture Structural Equation Model

A. The Model

Consider a general finite mixture structural equation model with G components. (We shall use the words 'component' and 'group' interchangeably.) Let $\mathbf{w} = (w_1, \ldots, w_G)'$ denote the vector of mixing proportions such that $w_g > 0 (g = 1, \ldots, G)$ and $\sum_{g=1}^{G} w_g = 1$. Let $\mathbf{y} \mid g$ denote a vector of p endogenous observable random variables and $\mathbf{x} \mid g$ a vector of q exogenous observable variables for the g-th component.

Assume that the $\mathbf{y} \mid g$ and $\mathbf{x} \mid g$ variables measure the unobservable $(m \times 1)$ vector of endogenous variables, η^g, and the unobservable $(n \times 1)$ vector of exogenous variables, ξ^g, according to the following measurement model:

$$\mathbf{y} \mid g = \nu_y^g + \Lambda_y^g \eta^g + \varepsilon^g \ , \tag{1}$$

$$\mathbf{x} \mid g = \nu_x^g + \Lambda_x^g \xi^g + \delta^g \ , \tag{2}$$

where $\Lambda_y^g (p \times m)$ and $\Lambda_x^g (q \times n)$ are coefficient matrices (factor loadings), $\nu_y^g (p \times 1)$, and $\nu_x^g (q \times 1)$ are vectors of measurement intercept terms, and ε^g and δ^g are vectors of measurement errors in $\mathbf{y} \mid g$ and $\mathbf{x} \mid g$ respectively. Assume that the measurement errors are uncorrelated with the unobservable variables, and satisfy $E(\varepsilon^g) = \mathbf{0}$, $E(\delta^g) = \mathbf{0}$, $E(\varepsilon^g \varepsilon^{g\prime}) = \Theta_\varepsilon^g$ and $E(\delta^g \delta^{g\prime}) = \Theta_\delta^g$ where Θ_ε^g and Θ_δ^g are not necessarily diagonal. Let $E(\xi^g) = \tau_\xi^g$ and $E[(\xi^g - \tau_\xi^g)(\xi^g - \tau_\xi^g)'] = \Phi^g$. Suppose the unobservable variables are related via the following system of linear structural relations:

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

where $\mathbf{B}_g$ is a $(m \times m)$ matrix of structural parameters specifying the links among the endogenous latent variables, Γ^g is a $(m \times n)$ coefficient matrix denoting the effect of ξ^g on η^g, α^g is a $(m \times 1)$ vector of intercept terms, and ζ^g $(m \times 1)$ is a random vector of disturbances (errors in equations). We assume that $E(\zeta^g \zeta^{g\prime}) = \Psi^g$, $E(\zeta^g) = \mathbf{0}$, ζ^g is uncorrelated with ξ^g, and $\mathbf{B}_g$ is nonsingular.[1]

---

1. We thank an anonymous referee for pointing out that one can express the structural equation model (see equations (1), (2) and (3)) parsimoniously using the two matrices specified in the Reticular Action Model [RAM] (McArdle 1978; McArdle and McDonald 1984). For a succinct discussion of the RAM, see McDonald (1985, pp. 151-156 and, in particular, equations (4.3.6) and (4.3.7)).

Given the above assumptions, the conditional mean vectors $\mu_g[(p+q)\times 1]$ and conditional covariance matrices $\Sigma_g[(p+q)\times(p+q)]$ of the joint vector $\Delta \mid g = \begin{bmatrix} \mathbf{y} \mid g \\ \mathbf{x} \mid g \end{bmatrix}$ are (see Jöreskog 1973):

$$\mu_g = \begin{bmatrix} \nu_y^g + \Lambda_y^g \mathbf{B}_g^{-1} (\alpha^g + \Gamma^g \tau_\xi^g) \\ \nu_x^g + \Lambda_x^g \tau_\xi^g \end{bmatrix}, \tag{4}$$

and

$$\Sigma_g = \begin{bmatrix} \Lambda_y^g \mathbf{B}_g^{-1} (\Gamma^g \Phi^g \Gamma^{g\prime} + \Psi^g) \mathbf{B}_g^{-1\prime} \Lambda_y^{g\prime} + \Theta_\varepsilon^g & \Lambda_y^g \mathbf{B}_g^{-1} \Gamma^g \Phi^g \Lambda_x^{g\prime} \\ \Lambda_x^g \Theta^g \Gamma^{g\prime} \mathbf{B}_g^{-1\prime} \Lambda_y^{g\prime} & \Lambda_x^g \Phi^g \Lambda_x^{g\prime} + \Theta_\delta^g \end{bmatrix}. \tag{5}$$

Before estimating the finite mixture structural equation model (i.e., the parameters of equations (4) and (5) and $\mathbf{w}$), we need to show that the model is identified. In contrast to the standard multigroup model with known group membership, we need to impose distributional assumptions to achieve model identification. Suppose the data follow a finite mixture of multivariate normal distributions. Then the finite mixture structural equation model is identified if the multigroup structural equation model for known group membership is identified. This result follows trivially because finite mixtures of multivariate normals are identified in the sense of Definition 3.1.1 in Titterington et al. (1985, p. 36). Henceforth we shall assume that the finite mixture structural equation model is identified.

Assume that $\Delta \mid g$ follows a conditional multivariate normal distribution with mean vector μ_g and covariance matrix Σ_g. Then the unconditional distribution of the observed vector $\Delta = \begin{bmatrix} \mathbf{y} \\ \mathbf{x} \end{bmatrix}$ is

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

where $f_g(\cdot)$ is the conditional multivariate normal density function:

$$f_g(\Delta \mid \mu_g, \Sigma_g) = (2\pi)^{-(\frac{p+q}{2})} \mid \Sigma_g \mid^{-1/2}$$

$$\exp\{-\frac{1}{2}(\Delta - \mu_g)' \Sigma_g^{-1} (\Delta - \mu_g)\}. \tag{7}$$

The unconditional likelihood function for a sample $(\Delta_1, \ldots, \Delta_N)$ of N randomly drawn observations from the finite mixture is then:

$$L = \prod_{i=1}^{N} [\sum_{g=1}^{G} w_g (2\pi)^{-(\frac{p+q}{2})} |\Sigma_g|^{-1/2}$$
$$\exp\{-1/2(\Delta_i - \mu_g)'\Sigma_g^{-1}(\Delta_i - \mu_g)\}], \quad (8)$$

where L is a function of the elements of $\mathbf{B}_g$, 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 prespecified number of groups G, while taking into account the constraints imposed on $\mathbf{w}$ above, and $|\Sigma_g| > 0$ for all g. The condition $|\Sigma_g| > 0$ is necessary because consistent estimators are not possible when Σ_g is singular.

Applying Bayes' rule, we can use (within any estimation iteration) the maximum likelihood estimates $\hat{\Sigma}_g$ and $\hat{\mu}_g$ (which are explicit functions of the estimated model parameters) and $\hat{w}_g$ to estimate the posterior probabilities, $\hat{P}_{ig}$, of each observation i in each of the G components:

$$\hat{P}_{ig} = \frac{\hat{w}_g f_g(\Delta_i \mid \hat{\Sigma}_g, \hat{\mu}_g)}{\sum_{k=1}^{G} \hat{w}_k f_k(\Delta_i \mid \hat{\Sigma}_k, \hat{\mu}_k)}. \quad (9)$$

These posterior probabilities represent a fuzzy classification (clustering) of the N observations into G components based on the postulated measurement and structural models.

B. Special Cases

The finite mixture structural equation model described by equations (1), (2), and (3) is very general and subsumes a variety of models as special cases. Reducing the model to either equation (1) or (2) produces a finite mixture confirmatory factor analysis model. If the endogenous and exogenous variables (factors) are measured by error-free indicators (i.e., $\nu_y^g = \mathbf{0}$, $\nu_x^g = \mathbf{0}$, $\Lambda_x^g = \mathbf{I}$, $\Lambda_y^g = \mathbf{I}$, $\Theta_\delta^g = \mathbf{0}$, and $\Theta_\varepsilon^g = \mathbf{0}$) the general model reduces to equation (3): a finite mixture, simultaneous equation model. We can impose various constraints on the general model parameters to produce a series of nested models. For example, we can make the measurement model parameters Λ^g and ν^g invariant across groups if it is reasonable to assume that the groups react similarly to the measuring instruments for ξ and η. Likewise, we can impose invariance across groups on the covariance matrices of errors. We can impose several other model restrictions depending on the context being studied or the theory being tested. Finally, the general structural equation mixture model is equivalent to multiple-group structural equation modeling (Jöreskog 1971; Sörbom 1974) if the groups and associated group memberships are *known* a priori.

C. Model Estimation: The EM Algorithm

This section outlines the full-information algorithm to estimate the general structural equation finite mixture model[2] (STEMM). Technical details and algebraic derivations are available from the first author.

Suppose the structural equation finite mixture model is identified (see sub-Section 3.B) and, in particular, the data follow a finite mixture of multivariate normals. For an EM algorithm formulation (see Dempster, Laird, and Rubin 1977) we begin by defining indicator variables z_{ig} as:

$$z_{ig} = \begin{cases} 1 & \text{iff observation } i \text{ belongs to group } g \text{ , and} \\ 0 & \text{otherwise.} \end{cases}$$

We also assume that, for a particular observation i, the non-observed vector, $\mathbf{z}_i = (z_{i1}, \ldots, z_{iG})'$, is i.i.d. multinominally distributed with probability vector $\mathbf{w}$. That is,

$$(\mathbf{z}_i \mid \mathbf{w}) \sim \prod_{g=1}^{G} w_g^{z_{ig}} . \tag{10}$$

The distribution of Δ_i given $\mathbf{z}_i$ is therefore:

$$(\Delta_i \mid \mathbf{z}_i) \sim \sum_{g=1}^{G} z_{ig} f_g(\Delta_i \mid \mu_g, \Sigma_g) = \prod_{g=1}^{G} [f_g(\Delta_i \mid \mu_g, \Sigma_g)]^{z_{ig}} . \tag{11}$$

With $\mathbf{Z} = ((z_{ig}))$ considered as missing data and $\Delta = ((\Delta_{ij}))$ as the input data matrix, the complete log-likelihood function to be maximized is given by:

---

2. An alternative strategy is to use a two-step limited-information estimator. In the first step, estimate μ_g, Σ_g and w_g (see equations (4) and (5) using a standard algorithm for estimating *unconstrained* mixtures of multivariate normals (e.g., Wolfe's NORMIX algorithm 1967)). In the second step, apply the standard multigroup methodology to the partitioned data (see Jöreskog 1971; Sörbom 1974). The two-step limited-information estimator is straightforward and easy to implement. However, the full-information estimator is more efficient for correctly specified mixture models which are overidentified (as is usually the case). See Judge, Griffiths, Hill, and Lutkepohl (1982, pp. 384-386) for a discussion. Furthermore, the two-step estimator is likely to encounter singularity problems in the first step if μ_g and Σ_g have high dimensionalities and the sample size is not sufficiently large. Note that the full-information and limited-information estimators are identical if the structural model for known group membership is exactly identified (e.g., the case of a finite mixture, single factor, three-indicator model where all the parameters vary across groups). In such a case, we recommend the simple two-step method because the M-step of the EM algorithm (Dempster, Laird, and Rubin 1977) exists in closed form.

$$\ln L_c(\mu,\Sigma,\mathbf{w} \mid \Delta,\mathbf{Z}) = -\left(\frac{p+q}{2}\right) \ln (2\pi) \sum_{i=1}^{N} \sum_{g=1}^{G} z_{ig}$$
$$- 1/2 \sum_{i=1}^{N} \sum_{g=1}^{G} Z_{ig} \ln |\Sigma_g|$$
$$- 1/2 \sum_{i=1}^{N} \sum_{g=1}^{G} z_{ig} (\Delta_i - \mu_g)' \Sigma_g^{-1} (\Delta_i - \mu_g)$$
$$+ \sum_{i=1}^{N} \sum_{g=1}^{G} z_{ig} \ln w_g \,, \tag{12}$$

where $\mu = (\mu_1, \ldots, \mu_G)$ and $\Sigma = (\Sigma_1, \ldots, \Sigma_G)$.

Note that the EM algorithm provides monotone increasing values of the log-likelihood function so that convergence to at least a locally optimum solution can be achieved (see Titterington, Smith, and Makov 1985, pp. 95-97). However, EM algorithms tend to converge slowly (see McLachlan and Basford 1988, p. 17). This issue will be addressed in Section 4.

The EM algorithm requires two steps in maximizing equation (12): an E-step in which we compute the expected value of $\mathbf{z}_i$ given Δ and provisional estimates for μ, Σ, and $\mathbf{w}$, and an M-step where we use Powell's (1977) conjugate gradient method to estimate $\mathbf{B} = (\mathbf{B}_1, \ldots, \mathbf{B}_G)$, $\Gamma = (\Gamma^1, \ldots, \Gamma^G)$, $\Lambda_x = (\Lambda_x^1, \ldots, \Lambda_x^G)$ $\Lambda_y = (\Lambda_y^1, \ldots, \Lambda_y^G)$, $\Theta_\delta = (\Theta_\delta^1, \ldots, \Theta_\delta^G)$, $\Theta_\varepsilon = (\Theta_\varepsilon^1, \ldots, \Theta_\varepsilon^G)$, $\Psi = (\Psi^1, \ldots, \Psi^G)$, $\Phi = (\Phi^1, \ldots, \Phi^G)$, $\nu_x = (\nu_x^1, \ldots, \nu_x^G)$, $\nu_y = (\nu_y^1, \ldots, \nu_y^G)$, $\alpha = (\alpha^1, \ldots, \alpha^G)$, $\tau_\xi = \tau_\xi^1, \ldots, \tau_\xi^G)$, and $\mathbf{w}$ conditional on the newly estimated values of $\mathbf{Z}$ and subject to the model identification restrictions. We iterate between the E- and M-steps until no further improvement in the log-likelihood function in equation (12) is possible.

1. E-Step

Using Bayes' theorem, we can show that (see Jedidi, Ramaswamy, and DeSarbo 1993, for a similar derivation):

$$E(z_{ig} \mid \Delta_i, \hat{\Sigma}, \hat{\mu}, \hat{\mathbf{w}}) = \frac{\hat{w}_g f_g(\Delta_i \mid \hat{\Sigma}_g, \hat{\mu}_g)}{\sum_{k=1}^{G} \hat{\mathbf{w}}_k f_k(\Delta_i \mid \hat{\Sigma}_k, \hat{\mu}_k)} \,, \tag{13}$$

which is equal to $\hat{P}_{ig}$ in equation (9). Therefore, the conditional expectation of equation (12) with respect to $\mathbf{z}_i$, evaluated at the provisional estimates $\hat{\Sigma}$, $\hat{\mu}$, $\hat{\mathbf{w}}$ (ignoring the constant term without loss of generality) is:

$$E_{\mathbf{z}}(\ln L_c \mid \Delta,\hat{\Sigma},\hat{\mu},\hat{\mathbf{w}}) = -\frac{1}{2}\sum_{g=1}^{G} N_g[\ln \mid \Sigma_g \mid + \operatorname{tr}(\mathbf{T}_g\Sigma_g^{-1}) - 2\ln w_g], \tag{14}$$

where

$$N_g = \sum_{i=1}^{N} \hat{P}_{ig},$$

$$\mathbf{T}_g = \frac{1}{N_g}\sum_{i=1}^{N} \hat{P}_{ig}(\Delta_i - \mu_g)(\Delta_i - \mu_g)'.$$

Thus, the expectation phase amounts to replacing the non-observed data $\mathbf{Z} = ((z_{ig}))$ in equation (12) by their estimated conditional expectations $\hat{\mathbf{P}} = ((P_{ig}))$.

2. M-Step

Let $F = -E_{\mathbf{z}}(\cdot)$ where $E_{\mathbf{z}}(\cdot)$ is equation (14). In this step, we minimize F with respect to $\mathbf{B}$, Γ, Λ_x, Λ_y, Θ_δ, Θ_ε, Ψ, Φ, ν_x, ν_y, α, τ_ξ and $\mathbf{w}$, subject to the identification restrictions and the constraints $w_g > 0$ and $\sum_{g=1}^{G} w_g = 1$, conditional on the new provisional estimates of $\mathbf{Z} = ((z_{ig}))$. To estimate $\mathbf{w}$, it suffices to minimize the augmented function

$$-\sum_{g=1}^{G} N_g \ln w_g + \theta\left(\sum_{g=1}^{G} w_g - 1\right), \tag{15}$$

where θ denotes a Lagrangian multiplier. By differentiating equation (15) with respect to θ and w_g and setting the derivatives equal to zero, we can show that:

$$\hat{w}_g = \frac{N_g}{N}. \tag{16}$$

Given $\hat{\mathbf{P}} = ((\hat{P}_{ig}))$ and $\hat{\mathbf{w}}$, we can minimize F with respect to the remaining parameters.

We estimate Λ_x, Λ_y, $\mathbf{B}$, Γ, Θ_δ, Θ_ε, Ψ, Φ, ν^g, τ_ξ^g, and α^g using Powell's (1977) conjugate gradient iterative method. We use the estimates from the M-Step to compute new values for $\hat{\mu}$ and $\hat{\Sigma}$ using equations (4) and (5). We use these new estimates for μ, Σ, and $\mathbf{w}$ in the next E-step iteration of the EM algorithm to update the posterior probabilities P_{ig} in equation (9). We alternate between the E- and M-steps until no further improvement in the likelihood function in equation (12) is possible. Once convergence occurs, we obtain final estimates for ν^g, τ_ξ^g, α^g, Λ_x^g, Λ_y^g, $\mathbf{B}_g$, Γ^g, Θ_δ^g, Θ_ε^g, Ψ^g, and

$\mathbf{P} = ((P_{ig}))$ for $g = 1, \ldots, G$ and compute the asymptotic standard errors of the model parameters using the inverse of the empirical information matrix (see Meilijson 1989).

D. Model Selection

In most practical cases, we expect the researcher to use the finite mixture structural equation model in a partially confirmatory mode because the number of components (groups) G is typically not known a priori. It is therefore necessary to choose statistical criteria for determining the optimal number of components (groups). In contrast to the standard multigroup case, conventional tests based on the likelihood ratio (e.g., X^2 statistic) do not apply because the regularity conditions are violated. See Titterington, Smith, and Makov (1995, pp. 4-5) for a simple example. This problem is widely recognized in the literature, and a variety of heuristics are available (see McLachlan and Basford 1988, pp. 35-36).

In keeping with the spirit of mixture models and structural equation modeling, we propose that the researcher examine several global measures of fit. One approach is to choose G to minimize the Akaike (1974) Information Criterion defined by:

$$\text{AIC}_G = -2 \ln L_G + 2M_G \tag{17}$$

where M_G denotes the number of free parameters estimated in a G-component solution (see Sclove 1983; Bozdogan and Sclove 1984). Note that M_G depends on the specific structural equation model being tested: there is no generic equation form for M_G. Bozdogan (1987) proposes modifying this AIC heuristic so that the penalty coefficient multiplying M_G is 3. We shall use this measure and refer to it is as the "modified AIC." A problem with the AIC criterion is that it fails to penalize overparametrization appropriately (see McDonald 1989). To deal with this deficiency, Bozdogan (1987) suggests that G be chosen to minimize the consistent AIC statistic defined by:

$$\text{CAIC}_G = -2 \ln L_G + M_G(\ln N + 1)\,. \tag{18}$$

Bozdogan asserts that CAIC is more robust than AIC, particularly in large samples (i.e., CAIC is less likely to lead to overparameterization).

Another approach is to use the Bayesian information criterion (Schwarz 1978) defined by:

$$\text{BIC} = -2 \ln L_G + M_G \ln N\,. \tag{19}$$

BIC is very similar to CAIC in penalizing overparametrization. We remind the reader that AIC, CAIC, and BIC are heuristics. (See Windham and Cutler

1992 for recent comparisons of methods for selecting G in finite mixture models.)

Assuming that a G-component model is satisfactory, it is useful to assess the classification of subjects into groups. We use an entropy-based measure E_G which is based on the posterior probabilities (Ramaswamy, DeSarbo, Reibstein, and Robinson 1992). Specifically, E_G is defined by:

$$E_g = 1 - [\sum_i \sum_g -\hat{P}_{ig} \ln \hat{P}_{ig}] / (N \ln G) . \quad (20)$$

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).

Finally, if one chooses to use a fully confirmatory model (i.e., one is willing to specify the number of components a priori or impose additional structure on the models applying to different groups), one can use more stringent statistical tests for model adequacy. For example, if G is known a priori, one can compare alternative nested models using the likelihood ratio statistic directly since the regularity conditions hold.

4. Testing the Mixture Model: Simulation Evidence

We performed simulations to test the EM mixture algorithm (STEMM) for three commonly-used model structures: (a) a confirmatory factor model, (b) a non-recursive model with feedback, and (c) a recursive model. Each model includes a fairly large number of parameters, ranging from 28 for the confirmatory factor model to 30 for the recursive and non-recursive models. To approximate real-life applications, we used a moderate random sample of 400 in all our simulations. For each model structure, we used the implied population mean vectors and covariance matrices for each group to generate a random sample. We used each sample to estimate the relevant model 100 times, using different sets of random starting values for the parameters. We used a convergence criterion of 0.00001 for improvements in the log-likelihood function and set the maximum number of iterations at 2,000. We paid particular attention to parameter recovery, local optima, convergence speed, and the performance of different model selection criteria. We performed all the runs on an IBM-RS6000 mainframe computer.

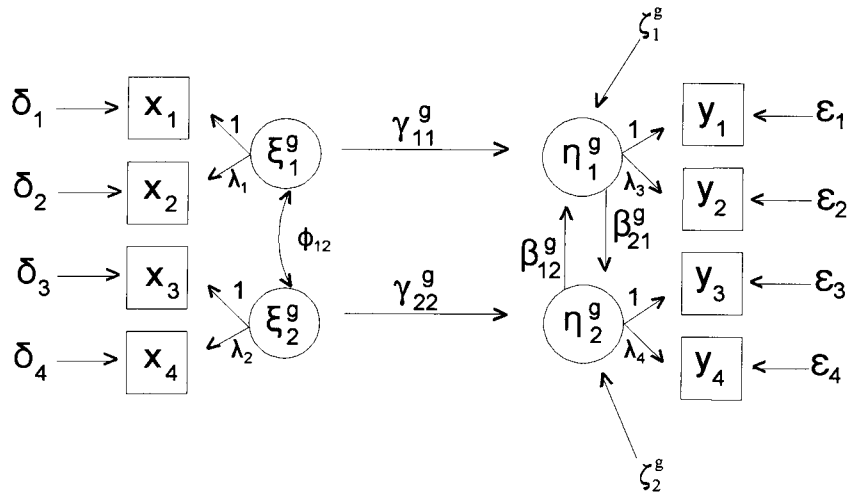
The first simulation analyzes a two-group confirmatory factor model with four factors, each measured by two indicators. Both groups are of equal size and follow the same model structure. However, the factor means and the reliabilities of the measures differ across groups. The population parameter values are shown in Table 1, column 2.

Table 1

TWO-GROUP SOLUTION: SIMULATED DATA FOR FOUR-FACTOR CONFIRMATORY MODEL

	Population Parameter	Estimated Parameter	Estimation Error
Measurement Model*			
x_2	1.000	1.063	0.063
x_4^1	1.000	0.977	-0.023
x_6^1	1.000	0.996	-0.004
x_8^1	1.000	1.068	0.068
x_2^2	1.500	1.536	0.036
x_4^2	1.500	1.476	-0.024
x_6^2	1.500	1.497	-0.003
x_8^2	1.500	1.505	0.005
Measurement Error			
θ_{11}	0.500	0.485	-0.015
θ_{22}	0.500	0.550	0.050
θ_{33}	0.500	0.497	-0.003
θ_{44}	0.500	0.469	0.031
θ_{55}	0.500	0.551	0.051
θ_{66}	0.500	0.520	0.020
θ_{77}	0.500	0.552	0.052
θ_{88}	0.500	0.533	0.033
Factor Structure			
ϕ_{11}	1.000	0.865	-0.135
ϕ_{12}	0.300	0.241	-0.059
ϕ_{13}	0.300	0.285	-0.015
ϕ_{14}	0.300	0.242	-0.058
ϕ_{22}	1.000	0.913	-0.087
ϕ_{23}	0.300	0.297	-0.003
ϕ_{24}	0.300	0.233	-0.067
ϕ_{33}	1.000	1.011	0.011
ϕ_{34}	0.300	0.277	-0.023
ϕ_{44}	1.000	0.820	-0.180
Factor Means			
$E(\xi_1^1 - \xi_1^2)$	3.00	2.83	-0.170
$E(\xi_2^1 - \xi_2^2)$	3.00	2.97	-0.030
$E(\xi_3^1 - \xi_3^2)$	3.00	2.98	-0.020
$E(\xi_4^1 - \xi_4^2)$	3.00	3.04	0.040
Mixing Proportion			
w_1	0.500	0.5027	0.003

*The parameters for x_1^g, x_3^g, x_5^g and x_7^g were set to unity for model identification.
* Superscripts denote groups.



Note: Superscripts are used for the groups (g = 1, 2) only when parameters differ across groups.

Figure 1. Theoretical Model for Simulated Data: Two-Group Non-Recursive Model.

Table 2

SUMMARY STATISTICS FOR MODEL SELECTION: SIMULATED DATA FOR FOUR-FACTOR CONFIRMATORY MODEL

G	-Ln L	AIC	CAIC	BIC	GFI	ITER
1	5231.6	10529.2	10617.0	10595.0	0.993	1953
2	5076.3	10245.7*	10369.4*	10338.4*	0.976	306
3	5068.8	10257.6	10417.2	10377.2	0.967	2205
4	5068.7	10284.4	10480.1	10431.1	0.968	2229

*Denotes minimum values for AIC, CAIC, and BIC.

We generated a random sample of 400, assuming that the data for each group follow a multivariate normal distribution. We then estimated the two-

group confirmatory factor model 100 times using different sets of random starting values for the parameters. The average CPU time for the two-group solution was 53.67 seconds, with a range of 34 to 122 seconds. The average number of iterations was 305.82, and the range was 259 to 375 iterations. Titterington, Smith, and Makov (1985, p. 89) report similar results with the EM algorithm. Of the 100 runs, 91 converged successfully to the largest stationary maximum (found via simulation). The remaining runs converged to local optima.

Table 2 provides the average summary statistics for model selection based on the successful runs. The information-theoretic criteria (i.e., the modified AIC, CAIC, and BIC) correctly point to the two-group model. The GFI criterion does not reflect the heterogeneity in the population. This result is consistent with previous findings that GFI is an inappropriate goodness-of-fit statistic even if the data belong to one population (see Mulaik, James, Alstine, Bennett, Lind, and Stilwell 1989).

Table 1 shows the parameter estimates for the two-group mixture model, averaged over the 91 successful runs. (Note that the parameter estimates were highly similar, typically differing only in the second decimal place.) The results show that the algorithm almost perfectly recovers the mixing proportions ($\hat{w}_1 = 0.5027$, $w_1 = 0.5$). In general, the parameter estimates are highly satisfactory, especially for the measurement model and the differences in factor means. Ignoring trivial fits for fixed parameters, the mean absolute percentage differences for the parameters of the measurement model and the factor mean structure are respectively 2.54% and 2.17%.

The second simulation analyzes a non-recursive, two-group model with two endogenous and two exogenous constructs, each with two indicators (see Figure 1). The structural parameters and factor means differ by group. However the measurement model is invariant across groups. The population parameter values are shown in Table 3, column 2.

We generated a random sample of 400, assuming that the data for each group follow a multivariate normal distribution. Given the complexity of the model structure (note that there was no structural model in the first simulation), it is not surprising that the EM algorithm required significantly more CPU time and iterations than in the first study. The average CPU time for the 100 runs was 384.12 seconds, and the CPU times ranged from 227 to 786 seconds. The average number of iterations per run was 1242.68, and the range was from 874 to 1601 iterations. Of the 100 runs, 74 converged successfully to the largest stationary maximum (found via simulation) and 26 converged to local optima. The parameter estimates for the successful runs were highly similar, typically differing only in the second decimal place.

Table 4 provides the average summary statistics for model selection based on the successful runs. As in the first simulation, the GFI criterion

Table 3
TWO-GROUP SOLUTION: SIMULATED DATA
FOR NON-RECURSIVE MODEL

	Population Parameter	Estimated Parameter	Estimation Error
Measurement Model: (Exogenous Variables)			
x_1	1	1*	
x_2	1	1.033	0.033
x_3	1	1*	
x_4	1	1.028	0.028
Measurement Model: (Endogenous Variables)			
y_1	1	1*	
y_2	1	0.97	-0.03
y_3	1	1*	
y_4	1	0.89	-0.11
Measurement Error (Exogenous Variables)			
θ_{11}	0.500	0.564	0.064
θ_{22}	0.500	0.370	-0.130
θ_{33}	0.500	0.645	0.145
θ_{44}	0.500	0.588	0.088
Measurement Error (Endogenous Variables)			
ϵ_{11}	0.500	0.41	-0.090
ϵ_{22}	0.500	0.47	-0.030
ϵ_{33}	0.500	0.33	-0.170
ϵ_{44}	0.500	0.53	0.030
Covariances (Exogenous Constructs)			
ϕ_{11}	1.00	1.067	0.067
ϕ_{12}	0.30	0.477	0.177
ϕ_{22}	1.00	1.008	0.008
Structural Parameters**			
β_{12}^1	-0.300	-0.449	-0.149
β_{21}^1	0.700	0.967	0.267
β_{12}^2	0.700	0.896	0.196
β_{21}^2	-0.300	-0.269	0.031
γ_{11}^1	0.500	0.592	0.092
γ_{22}^1	0.500	0.561	0.061
γ_{11}^2	-0.500	-0.500	0.000
γ_{22}^2	-0.500	-0.677	-0.177
Structural Errors			
Ψ_{11}	0.500	0.66	0.160
Ψ_{12}	0.000	0.11	0.110
Ψ_{22}	0.500	0.60	0.100
Factor Means			
$E(\xi_1^1 - \xi_1^2)$	3.00	2.81	-0.190
$E(\xi_2^1 - \xi_2^2)$	3.00	2.89	-0.110
Mixing Proportion			
w_1	0.5	0.494	-0.006

*denotes parameters set to unity for model identification.
**Superscripts denote groups.

Table 4

SUMMARY STATISTICS FOR MODEL SELECTION: SIMULATED DATA FOR NON-RECURSIVE MODEL

G	-Ln L	AIC	CAIC	BIC	GFI	ITER
1	4804.3	9674.6	9762.4	9740.4	0.993	2324
2	4703.9	9500.8	9624.5*	9593.5*	0.973	1500
3	4690.1	9500.3*	9659.9	9619.9	0.971	2420
4	4687.5	9522.0	9717.6	9668.6	0.966	3885

*Denotes minimum values for AIC, CAIC, and BIC.

leads to underfitting (i.e., the one-group solution). In contrast to the first study, the modified AIC criterion leads to overfitting and points to the three-group solution (see also Cudeck and Browne 1983). However, the CAIC and BIC statistics point to the two-group solution.

Table 3 gives the parameter estimates for the two-group solution, averaged over the 74 successful runs. The algorithm almost perfectly recovers the mixing proportions ($\hat{w}_1 = 0.4936$, $w_1 = 0.5$). Overall, the parameter estimates are satisfactory, especially for the measurement model and the differences in factor means. The mean absolute percentage differences (ignoring trivial fits) for the parameters of the measurement model and factor mean structure are respectively 5.03% and 5.00%.

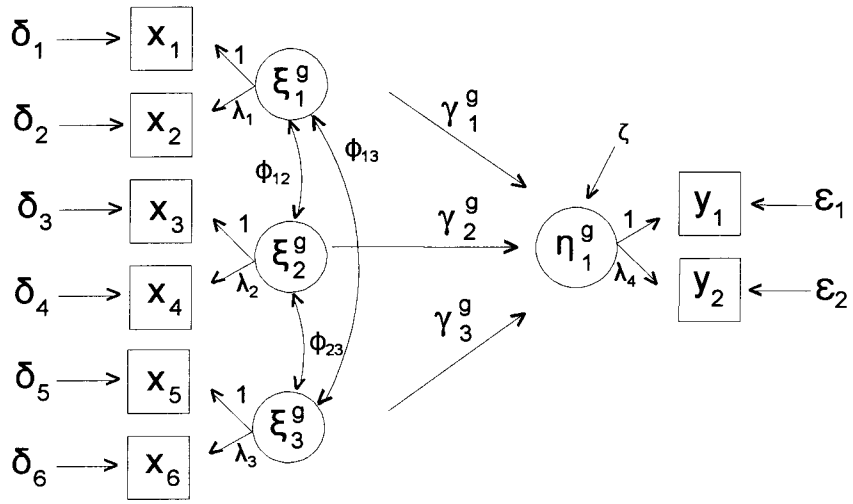
The third simulation analyzes a recursive two-group model with one endogenous and three exogenous constructs, each with two measures (see Figure 2). Both groups have a common measurement model. However, the structural parameters and factor means differ by group. The population parameter values are shown in Table 5, column 2.

As in the previous simulations, we generated a random sample of 400, assuming that the data for each group have a multivariate normal distribution. Not surprisingly, the average CPU times and numbers of iterations for the two-group solution were larger than the corresponding values for the pure measurement model (first study) and smaller than the corresponding values for the non-recursive model (second study). The average CPU time was 116.36 seconds, and the range of CPU times was 52 to 290 seconds. The average number of iterations per run was 304.82, with a range of 219 to 428. Of the 100 runs, 85 converged successfully to the largest stationary maximum

Table 5
TWO-GROUP SOLUTION: SIMULATED DATA FOR RECURSIVE MODEL

	Population Parameter	Estimated Parameter	Estimation Error
Measurement Model:			
(Exogenous Variables)			
x_1	1.000	1.000*	
x_2	1.000	1.026	0.026
x_3	1.000	1.000*	
x_4	1.000	1.031	0.031
x_5	1.000	1.000	
x_6	1.000	0.984	-0.016
Measurement Model:			
(Endogenous Variables)			
y_1	1.000	1.000*	
y_2	1.000	0.990	-0.010
Measurement Error			
(Exogenous Variables)			
θ_{11}	0.500	0.515	0.015
θ_{22}	0.500	0.591	0.091
θ_{33}	0.500	0.559	0.059
θ_{44}	0.500	0.366	-0.134
θ_{55}	0.500	0.542	0.042
θ_{66}	0.500	0.489	-0.011
Measurement Error			
(Endogenous Variables)			
ϵ_{11}	0.500	0.340	-0.160
ϵ_{22}	0.500	0.490	-0.010
Covariances			
(Exogenous Constructs)			
ϕ_{11}			
ϕ_{12}	1.000	1.044	0.044
ϕ_{13}	0.300	0.455	0.155
ϕ_{22}	0.300	0.279	-0.021
ϕ_{23}	1.000	1.059	0.059
ϕ_{33}	0.300	0.286	-0.014
	1.000	0.980	-0.020
Structural Parameters**			
γ_1^1	0.500	0.436	-0.064
γ_2^1	0.500	0.468	-0.032
γ_3^1	0.500	0.614	0.114
γ_1^2	-0.500	-0.414	0.086
γ_2^2	-0.500	-0.633	-0.133
γ_3^2	-0.500	-0.456	0.044
Structural Errors			
Ψ_{11}	0.500	0.520	0.020
Factor Means			
$E(\xi_1^1 - \xi_1^2)$	3.000	2.960	-0.040
$E(\xi_2^1 - \xi_2^2)$	3.000	2.850	-0.150
$E(\xi_3^1 - \xi_3^2)$	3.000	2.980	-0.020
Mixing Proportion			
w_1	0.500	0.507	0.007

*denotes parameters set to unity for model identification.



Note: Superscripts are used for the groups (g = 1, 2) only when parameters differ across groups.

Figure 2. Theoretical Model for Simulated Data: Two-Group Recursive Model.

Table 6

SUMMARY STATISTICS FOR MODEL SELECTION: SIMULATED DATA FOR RECURSIVE MODEL

G	-Ln L	AIC	CAIC	BIC	GFI	ITER
1	5150.7	10367.5	10455.3	10433.2	0.993	1645
2	4873.2	9836.3*	9956.1*	9926.0*	0.967	346
3	4863.3	9840.6	9992.3	9954.2	0.962	1648
4	4862.3	9862.5	10046.1	10000.1	0.962	1239

*Denotes minimum values for AIC, CAIC, and BIC.

(found via simulation), and 15 converged to local optima. The parameter estimates for the successful runs were highly similar, typically differing only in the second decimal place.

Tables 5 and 6 respectively provide the average parameter values and fit statistics for model selection based on the successful runs.

All the information-based measures of fit (i.e., modified AIC, CAIC, and BIC) point correctly to the two-group solution. GFI leads to underfitting as in the second study. The algorithm almost perfectly recovers the mixing proportions ($\hat{w}_1 = 0.5068$, $w_1 = 0.5$) and the parameter estimates are satisfactory. In particular, the mean absolute percentage errors for the measurement parameters and the factor mean structures are respectively 2.08% and 2.33%.

These simulation experiments lead to several important conclusions. Only the CAIC and BIC measures of fit consistently pick the correct number of components (groups). The modified AIC works well in general but can lead to overfitting (i.e., choosing too many groups). GFI is too conservative and always leads to an underfitted model. Our algorithm (STEMM) provides satisfactory estimates even for reasonably large models and a moderate sample size. Although we did not encounter any problems stemming from empirical underidentification, local optima can, as in other iterative multivariate algorithms, be a problem especially when the sample size is moderate. Hence the preferred strategy is to estimate the model using different sets of starting values to identify multiple local optima if they exist. Finally, the simulation results reported above are limited, and future research should further test the robustness of the algorithm under different scenarios (e.g., different model structures).

5. An Empirical Application

Bagozzi, Baumgartner, and Pieters (1995) developed a second-order factor model to examine the role of emotion in goal-directed behavior (see Figure 3). Their model is based on a general framework for an emotional goal system in which an individual's appraisals of the consequences of achieving success/failure at a goal elicit anticipatory emotions. Bagozzi et al. (1995) postulate that "emotion" is a superordinate factor and "positive affect," "negative affect," and "volition" are first-order factors. Each first-order factor is measured using two or more observed subordinate indicators (y_1 through y_{10} in Figure 3). Each subordinate indicator for "positive affect" and "negative affect" is measured as the average score of two or more items, each of which is measured on an 11-point scale with anchor points "not at all" and "very much." For example, y_2 is the average score of three items: happy, glad, and satisfied. Each subordinate indicator for "volition" is measured using a single item with anchor points "completely agree" and "completely disagree."

Bagozzi et al. (1995) tested their second-order factor model in the context of regulating one's bodyweight by exercising and dieting. The sample consisted of 763 subjects in the Netherlands who were part of an ongoing panel of representatives, administered by the Netherlands' Institute for Public Opinion.

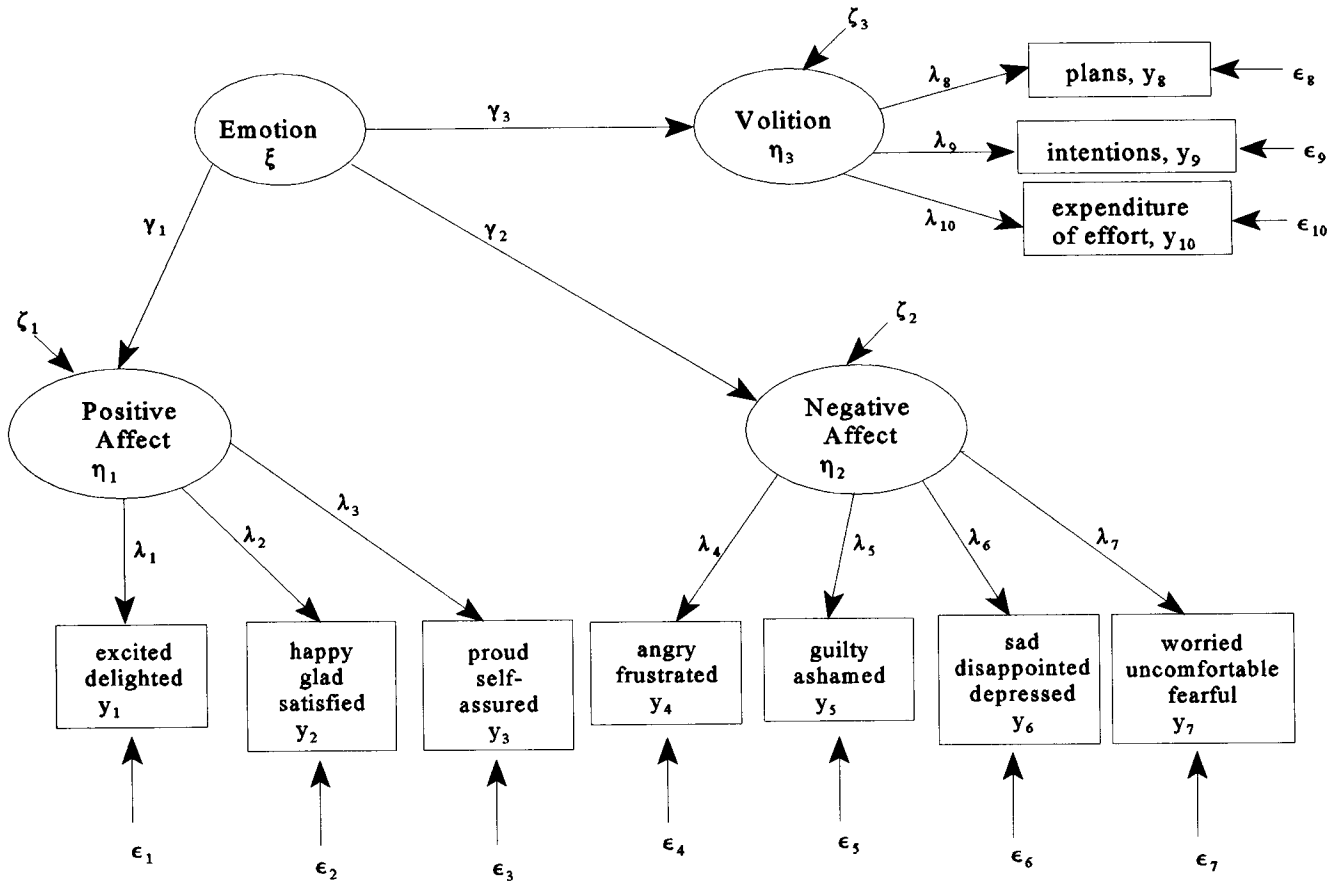


Figure 3. A Structural Equation Model of Emotion.

Table 7

SUMMARY STATISTICS FOR MODEL SELECTION IN THE EMOTION STUDY

G	Factor Loadings	M	-Ln L	CAIC	BIC	Adj GFI	E
1	-	23	10837.9	21839.9	21816.9	0.94	-
2	Invariant	51	10428.8	21221.5	21170.5	0.92	0.96
	Free	58	10414.3	21242.4	21184.4	0.93	0.96
3	Invariant	69	10296.2	21084.7*	21015.7*	0.95	0.76
	Free	84	10270.2	21132.6	21049.6	0.95	0.81
4	Invariant	87	10281.2	21183.2	21096.2	0.94	0.68
	Free	108	10247.3	21265.2	21157.2	0.95	0.73

*Denotes minimum values for CAIC and BIC.

Table 8

FACTOR MEAN SCORES FOR THE MIXTURE MODEL IN THE EMOTION STUDY

Factor	Group 1	Group 2	Group 3
Positive Affect	0	4.51	2.1
Negative Affect	0	3.15	1.45
Volition	0	3.44	1.02
Emotion	0	3.21	1.02

If the data were homogeneous, one could use the aggregate sample covariance matrix to estimate the structural equation model. In reality, different groups could follow different models of emotion-directed behavior. The structural parameters which link the superordinate factor "emotion" and the first-order factors (i.e., γ_1, γ_2, and γ_3) could vary across groups. For example, in the context of weight loss, "emotion" can affect intentional structures (e.g., "volition") differently for different groups of the population. Furthermore, the measurement model which links the first-order factors and the subordinate indicators (i.e., λ_1 through λ_{10} inclusive) could also differ across groups. For example, the effect of "volition" on the expenditure of effort (λ_{10}) can vary across people.

To test for heterogeneity, we used our algorithm (STEMM) to estimate the mixture model. Specifically, we used the partially confirmatory approach because extant theory cannot be used to determine the number of groups (components) a priori. We estimated seven models, varying the number of groups from 1 through 4 and allowing for both free and invariant first-order factor loadings (i.e., λ's) across groups (see Table 7). Both the CAIC and BIC criteria for model selection pointed to the three-group, invariant factor-loading model. We now analyze the results for this model.

Table 9

MEASUREMENT PARAMETERS FOR THE AGGREGATE AND MIXTURE SOLUTIONS IN THE EMOTION STUDY*

	Aggregate Solution		Mixture Solution			
				Error Variance		
Indicator	Loading	Error Variance	Loading	Group 1	Group 2	Group 3
λ_1	1	4.76	1	3.73	5.16	7.18
		(.42)		(.428)	(.411)	(.480)
λ_2	1.12	2.53	1.15	3.66	0.54	3.96
	(.05)	(.28)	(.07)	(.676)	(.093)	(.334)
λ_3	1.15	1.77	1.19	1.59	0.23	3.06
	(.05)	(.26)	(.07)	(.319)	(.084)	(.323)
λ_4	1	2.43	1	0.02	3.54	3.07
		(.23)		(.002)	(.216)	(.235)
λ_5	0.97	3.38	0.98	0.03	5.10	3.78
	(.03)	(.29)	(.006)	(.002)	(.262)	(.323)
λ_6	1.01	2.31	1.05	0.04	2.73	3.33
	(.03)	(.23)	(.009)	(.004)	(.249)	(.164)
λ_7	0.95	3.84	1.10	0.02	3.57	5.87
	(.03)	(.32)	(.006)	(.002)	(.318)	(.211)
λ_8	1	5.91	1	7.56	6.95	7.82
		(.52)		(.603)	(.535)	(.466)
λ_9	1.31	1.00	1.32	0.45	0.99	1.29
	(.06)	(.26)	(.09)	(.272)	(.168)	(.217)
λ_{10}	1.26	2.06	1.26	3.14	0.07	2.96
	(.06)	(.28)	(.08)	(.326)	(.134)	(.253)

*Standard errors in parentheses.

Table 10

STRUCTURAL PARAMETERS FOR THE AGGREGATE AND MIXTURE MODELS IN THE EMOTION STUDY*

Endogenous Factor	Aggregate Solution	Mixture Model		
		Group 1	Group 2	Group 3
Positive Affect (γ_1)	3.15	4.31	1.40	2.05
	(0.23)	(0.90)	(0.24)	(0.41)
Negative Affect (γ_2)	2.06	2.32	0.978	1.42
	(0.20)	(0.59)	(0.18)	(0.32)
Volition (γ_3)	1.75	1.66	1.07	0.99
	(0.17)	(0.42)	(0.18)	(0.24)
Mixing Proportions (w)	-	0.22	0.38	0.4

* Standard errors are in parentheses.

The three groups comprise 22%, 38%, and 40% respectively of the sample. The structured factor means for each group are shown in Table 8, where Group 1 denotes the reference group and also has the lowest means for the superordinate factor, "emotion," and all three first-order factors ("positive affect," "negative affect," and "volition"). We shall refer to Group 1 as the *low felt-emotion* group. Group 2 has the highest means for all four factors. Hence we refer to Group 2 as the *high felt-emotion* group and Group 3 as the *moderate felt-emotion* group.

The results suggest that the measurement model (i.e., the factor loadings for the subordinate indicators) is invariant across all three *felt-emotion* groups. Thus the emotion model exhibits a certain degree of generalizability across groups; see Table 9.

In contrast to the results for the measurement model, the data show that there is considerable heterogeneity in the structural model (i.e., the γ's vary across groups); see Table 10. The effect of "emotion" on any given first-order factor ("positive affect," "negative affect," and "volition") is the highest for Group 1 (the *low felt-emotion*) and the lowest for Group 2 (the *high felt-emotion*).

The next step in the analysis was to determine if group membership is related to demographic variables. For each group, we performed a logistic regression where $\log [P_{ig} / (1 - P_{ig})]$ is the dependent variable and gender (coded as 0 for males and 1 for females), household income, education, and the individual's weight-to-height ratio are the independent variables. We used the last variable because Bagozzi et al. (1995) examined individuals'

Table 11

LOGISTIC REGRESSION RESULTS FOR POSTERIOR ANALYSIS IN EMOTION STUDY[a]

Variable	Group 1	Group 2	Group 3
Intercept	-4.09	-8.82	3.77
Gender[b]	-2.06*	1.82**	0.35
Income	0.19*	-0.02	-0.19*
Education	-0.24	0.02	0.14
Weight	-0.38	1.34**	-0.75*

[a] All the regressions are significant at the 0.05 level.

[b] Coded as 0 for males and 1 for females.

** $p < 0.01$.

* $p < 0.05$.

emotional goals systems in the context of those individuals' regulation of their bodyweight.

Table 11 reports the results of the logistic regression. Three demographic variables (gender, income, and the weight-to-height ratio) are significantly related to group membership. Subjects in Group 1 (the *low felt-emotion*) are more likely to be males ($p < 0.05$) and have higher incomes ($p < 0.05$). Subjects in Group 2 (the *high felt-emotion*) are more likely to be females ($p < 0.01$) and have higher weight-to-height ratios ($p < 0.01$). In contrast, the subjects in Group 3 (the *moderate felt-emotion*) are more likely to have lower incomes ($p < 0.05$) and lower weight-to-height ratios ($p < 0.05$). Gender is not significant for this group.

These results suggest that females and subjects with higher weight-to-height ratios are more likely to experience higher felt-emotions than males and subjects with lower weight-to-height ratios. The analysis also indicates that low-income subjects experience more emotion than high- income subjects. Furthermore, it appears that these demographic groups follow different emotion models in regulating bodyweight.

Our finding on gender differences in the experienced intensity of anticipatory emotions is consistent with Bagozzi et al. (1995) and empirical studies of gender differences in emotion experiences, which find females to be more emotionally expressive than men (e.g., Wood, Rhodes, and Whelan 1989; Derbaix and Pham 1991). The result regarding higher-weight subjects is also consistent with Nisbett's (1972) finding that obese people are more emotional than their normal-weight counterparts. To the extent that income is related to self-esteem, our results seem consistent with previous research showing that self-esteem intensifies emotional reactions, especially in unpleasant settings (e.g., Weiner, Russel, and Lerman 1979; Brown and Dutton 1995).

6. Summary

This paper proposes a new structural equation finite mixture model and algorithm (STEMM) for simultaneously depicting heterogeneity and forming groups in the context of a postulated causal (i.e., simultaneous equation regression) structure in which there are measurement and structural errors. The model is very general and subsumes as special cases finite mixtures of confirmatory factor analysis, second-order factor analysis, and simultaneous equation models with feedback.

Our simulation results suggest that the CAIC and BIC fit statistics are reasonably robust for model selection. However, as in other multivariate methods (e.g., multigroup structural equation models with known group membership), the researcher should estimate the mixture model using a wide range of starting values to reduce the chance of obtaining local optima.

We analyzed a second-order factor model of emotion proposed by Bagozzi et al. (1995). The results suggest that different groups of the population follow different models of goal-directed behavior. For the data examined, gender, the weight-to-height ratio, and income were significant predictors of group membership. In particular, the results suggest that female and higher-weight (low-income) subjects experience higher *felt-emotions* than males and lower-weight (high-income) subjects.

Because of the computational burden, our simulation study was modest. Extensive and detailed simulation studies are necessary to analyze more complex model structures and to allow for different distributional forms (e.g., departures from multivariate normality). Furthermore, future research is necessary to develop better starting values, especially in cases where the number of components is unknown and a subsample of classified data is unavailable. Finally, an important area for future research is to reparameterize the mixing proportions as functions of such explanatory variables as demographics and psychographics to provide a richer explanation of heterogeneity (see Dayton and McReady 1988).

References

AITKIN, M., ANDERSON, D. and HINDE J. (1981), "Statistical Modeling of Data on Teaching Styles (with Discussion)," *Journal of the Royal Statistical Society, Series A, 144*, 419-461.

AKAIKE, H. (1974), "A New Look at Statistical Model Identification," *IEEE Transactions on Automatic Control, 6*, 716-723.

BAGOZZI, R. P., BAUMGARTNER, H., and PIETERS, R. (1995), "Goal-Directed Emotions," unpublished working paper, Ann Arbor: University of Michigan.

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.

BOZDOGAN, H. (1983), "Determining the Number of Component Clusters in Standard Multivariate Normal Mixture Models Using Model-Selection Criterion," *Technical Report *VIC/DQM/A83-1*, Army Research Office.

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

BOZDOGAN, H., and SCLOVE, S. L. (1984), "Multi-sample Cluster Analysis Using Akaike's Information Criterion," *Annals of the Institute of Statistical Mathematics, 36*, 163-180.

BROWN, J. D., and DUTTON, K. A. (1995), "The Thrill of Victory, the Complexity of Defeat: Self-Esteem and People's Emotional Reactions to Success and Failures," *Journal of Personality and Social Psychology, 68 (4)*, 712-722.

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

CUDECK, R., and BROWNE, M. W. (1983), "Cross-Validation of Covariance Structures," *Multivariate Behavioral Research, 18*, 147-167.

DAYTON, C. M., and MACREADY, G. B. (1988), "Concomitant-Variables Latent-Class Models," *Journal of the American Statistical Society, 83*, 173-178.

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

DERBAIX, C., and PHAM, M. T. (1991), "Affective Reactions to Consumption Situations," *Journal of Economic Psychology, 12*, 325-355.

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

JEDIDI, K., RAMASWAMY, V., and DESARBO, W. S. (1993), "A Maximum Likelihood Method for Latent Class Regression Involving a Censored Dependent Variable," *Psychometrika, 58*, 375-394.

JÖRESKOG, K. G.(1971), "Simultaneous Factor Analysis in Several Populations," *Psychometrika, 36*, 409-426.

JÖRESKOG, K. G.(1973), "A General Method for Estimating a Linear Structural Equation System," in *Structural Equation Models in the Social Sciences*, Ed., A. S. Goldberger and O. D. Duncan, New York: Academic Press, pp. 85-112.

JUDGE, C. G., GRIFFITHS, W. F., HILL, R. C., LUTKEPOHL, H., and LEE, T. C. (1982), *Introduction to the Theory and Practice of Econometrics*, New York: Wiley.

MCARDLE, J. J. (1978), "A Structural View of Structural Models," Paper Presented at Winter Workshop on Latent Structure Models Applied to Development Data, University of Denver.

MCARDLE, J. J. and MCDONALD, R. P. (1984), "Some Algebraic Properties of the Reticular Action Model for Moment Structures," *British Journal of Mathematical and Statistical Psychology, 37*, 234-251.

MCDONALD, R. P. (1985), *Factor Analysis and Related Methods*, New Jersey: Lawrence Erlbaum Associates.

MCDONALD, R. P. (1989), "An Index of Goodness-of-fit Based on Noncentrality," *Journal of Classification, 6*, 97-103.

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

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

MEILIJSON, I. (1989), "A Fast Improvement to the EM Algorithm on its Own Terms," *Journal of the Royal Statistical Society, 51*, 127-138.

MILLIGAN, G. W. (1980), "An Examination of the Effect of Six Types of Error Perturbation on Fifteen Clustering Algorithms," *Psychometrika, 45*, 325-342.

MILLIGAN, G. W. (1981), "A Review of Monte Carlo Tests of Cluster Analysis," *Multivariate Behavioral Research, 16*, 379-407.

MILLIGAN, G. W. and COOPER, M. C. (1987), "Methodology Review: Clustering Methods," *Applied Psychological Measurement, 11*, 4, 329-354.

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

MULAIK, S. A., JAMES, L. R., ALSTINE, V., BENNETT, J., LIND, N. S., and STILWELL, C. D. (1989), "Evaluation of Goodness-of-Fit Indices for Structural Equation Models," *Psychological Bulletin, 105 (May)*, 430-445.

MUTHÉN, B. O. (1989), "Latent Variable Modeling in Heterogeneous Populations," *Psychometrika, 54*, 557-585.

NISBETT, R. E. (1972), "Hunger, Obesity, and the Ventromedial Hypothalamus," *Psychological Review, 79*, 433-453.

POWELL, M. J. D. (1977), "Restart Procedures for the Conjugate Gradient Method," *Mathematical Programming, 12*, 241-254.

RAMASWAMY, V., DESARBO, W. S., REIBSTEIN, D. J, and ROBINSON, W. T. (1992), "An Empirical Pooling Approach for Estimating Marketing Mix Elasticities with PIMS Data," *Marketing Science, 12*, 241-254.

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

SCLOVE, S. L. (1983), "Application of the Conditional Population-Mixture Model to Image Segmentation," *IEEE Transactions on Pattern Analysis and Machine Intelligence, PAMI-5*, 428-433.

SÖRBOM, D. (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.

TITTERINGTON, D. M., SMITH, A. F. M., and MAKOV, V. E. (1985), *Statistical Analysis of Finite Mixture Distributions*, New York: Wiley.

WEINER, B., RUSSEL, D., and LERMAN, D. (1979), "The Cognition-Emotion Process in Achievement-Related Contexts," *Journal of Personality and Social Psychology, 37*, 1211-1220.

WINDHAM, M. P. and CUTLER, A. (1992), "Information Ratios for Validating Mixture Analyses," *Journal of the American Statistical Association, 87*, 1188-1192.

WOOD, W., RHODES N., and WHELAN, M. (1989), "Sex Differences in Positive Well Being: A Consideration of Emotional Style and Marital Status," *Psychological Bulletin, 35*, 165-175.

WOLFE, J. H. (1967), "NORMIX; Computational Methods for Estimating the Parameters of Multivariate Normal Mixtures of Distributions," Research Memorandum, SRM 68-2, U.S. Naval Personnel Research Activity, San Diego.